Horticulture
Research

Horticulture Research, 2022, **9**: uhac119

https://doi.org/10.1093/hr/uhac119

Article

Genome-wide core sets of SNP markers and Fluidigm assays for rapid and effective genotypic identification of Korean cultivars of lettuce (*Lactuca sativa* L.)

Jee-Soo Park[1], Min-Young Kang[1], Eun-Jo Shim[1], JongHee Oh[1], Kyoung-In Seo[1], Kyung Seok Kim[2], Sung-Chur Sim[3], Sang-Min Chung[4], Younghoon Park[5], Gung Pyo Lee[6], Won-Sik Lee[1], Minkyung Kim[3] and Jin-Kee Jung[1,*]

[1]Seed Testing and Research Center, Korea Seed & Variety Service, Gimcheon 39660, Republic of Korea
[2]Department of Natural Resource Ecology and Management, Iowa State University, Ames IA 50011, USA
[3]Department of Bioresources Engineering, Sejong University, Seoul 05006, Republic of Korea
[4]Department of Life Sciences, Dongguk University, Seoul 04620, Republic of Korea
[5]Department of Horticultural Bioscience, Pusan National University, Miryang 50463, South Korea
[6]Department of Plant Science and Technology, Chung-Ang University, Ansung 17546, South Korea
*Corresponding author. jinkeejung@korea.kr

Abstract

Lettuce is one of the economically important leaf vegetables and is cultivated mainly in temperate climate areas. Cultivar identification based on the distinctness, uniformity, and stability (DUS) test is a prerequisite for new cultivar registration. However, DUS testing based on morphological features is time-consuming, labor-intensive, and costly, and can also be influenced by environmental factors. Thus, molecular markers have also been used for the identification of genetic diversity as an effective, accurate, and stable method. Currently, genome-wide single nucleotide polymorphisms (SNPs) using next-generation sequencing technology are commonly applied in genetic research on diverse plant species. This study aimed to establish an effective and high-throughput cultivar identification system for lettuce using core sets of SNP markers developed by genotyping by sequencing (GBS). GBS identified 17 877 high-quality SNPs for 90 commercial lettuce cultivars. Genetic differentiation analyses based on the selected SNPs classified the lettuce cultivars into three main groups. Core sets of 192, 96, 48, and 24 markers were further selected and validated using the Fluidigm platform. Phylogenetic analyses based on all core sets of SNPs successfully discriminated individual cultivars that have been currently recognized. These core sets of SNP markers will support the construction of a DNA database of lettuce that can be useful for cultivar identification and purity testing, as well as DUS testing in the plant variety protection system. Additionally, this work will facilitate genetic research to improve breeding in lettuce.

Introduction

Lettuce (*L. sativa* L., $2n = 2x = 18$) is one of the agriculturally important leaf vegetables belonging to the Asteraceae family and is cultivated mainly in temperate climate areas of the world. In Korea, the production of lettuce was estimated to be over 93 543 tons from 3773 ha in 2018 [1]. The genome of lettuce has been completely decoded [2], and modeling analysis of approximately 45 000 genes has been conducted. New cultivars of lettuce are developed and commercialized every year, and identification of each cultivar is important for the registration and protection measures of newer cultivars. However, the identification of lettuce cultivars is a difficult task due to their close genetic relationships. Lettuce is a morphologically diverse crop and can be classified into different horticultural types based on head and leaf shape, size, structure, and stem length as well as end uses [3]. According to the International Union for the Protection of New Varieties of Plants (UPOV) TG/13/10 guidelines for lettuce [4], lettuce can be grouped into diverse types, such as butterhead, Iceberg, Frisée d'Amerique (loose-leaf), Oakleaf, and Cos (Romaine).

The UPOV defines the rights of plant breeders and protects them from unauthorized utilization of new cultivars. For cultivar registration and protection, distinctness, uniformity, and stability (DUS) testing in a UPOV system of plant variety protection (PVP) is required. However, DUS testing, which is based on morphological features, is costly, time-consuming, and labor-intensive. Furthermore, it can be influenced by various environmental conditions that can impose limitations on the identification of cultivars [5]. Therefore, DUS testing needs to be supported by genetic analysis using molecular markers. The UPOV has agreed to the deployment of

Received: 7 July 2021; **Accepted:** 4 May 2022; **Published:** 26 May 2022; **Corrected and Typeset:** 1 August 2022

molecular markers for the identification of cultivars that are specifically linked to a phenotypic trait [6]. In addition, the working group on Biochemical and Molecular Techniques and DNA-profiling in Particular (BMT), a technical committee under the UPOV, has discussed the usage of molecular markers for the identification and protection of cultivars [6, 7]. Diverse molecular markers, such as simple sequence repeats (SSRs) and single nucleotide polymorphisms (SNPs), have been applied in various plants for the identification and purity assessment of cultivars [8–16].

Before registration of a new cultivar, DUS tests are carried out to determine whether a new cultivar is distinct, uniform and stable. As a part of DUS testing, the distinctness of the new cultivar is examined by comparison with a similar existing cultivar, which is called "reference varietiy". SNP markers have been routinely utilized as a tool for the management of reference varieties for DUS examinations. Among the grouped cultivars, the one with the highest genetic similarity to the unknown cultivar can be selected as a "similar variety" and used for the DUS test. In other words, more relevant reference varieties can be selected for DUS testing based on their DNA profiles, and the duration and cost of DUS testing can be reduced.

SSR markers have been widely used for the assessment of phylogenetic relationships and DUS testing in commercial lettuce cultivars [10, 17–23] because of their advantages of being co-dominant and multi-allelic [24, 25] (Choi et al., 2016; Kong et al., 2020). Hong et al. [26] constructed expressed sequence tag-SSR profiles to identify 92 lettuce cultivars from Korea and proposed the utility of the markers in the distinctiveness tests of lettuce. Zhou et al. [8] developed a set of 19 SSR markers for the identification of 73 cultivars of head lettuce (*Lactuca L. sativa capitate* L.). SSR markers were also used to characterize the genetic diversity of the germplasm of chicory (*Cichorium intybus*), which also belongs to the same family, Asteraceae, as lettuce [27]. However, SSR markers have limitations since they are less reproducible and time-consuming and expensive to develop [28].

Compared to SSR markers, SNP markers are bi-allelic, making it simple to merge data between groups, and it is possible to generate large databases of marker information coupled with high-throughput genotyping. Currently, SNPs are the most preferred in cultivar identification and genomic studies due to their high abundance, stability, and efficiency [29, 30]. For example, a large collection of SNPs was identified from 223 pumpkin cultivars via genotyping by sequencing (GBS), and core markers were selected for cultivar identification in pumpkin [14]. Phylogenetic studies, evaluation of genetic variation and population structure, genome-wide association studies, and construction of genetic linkage maps based on lettuce SNP markers and genotype data have been applied to facilitate efficient genetic studies in lettuce [3, 31–34]. Truong et al. [31] constructed the linkage map of lettuce using 1113 SNPs via sequence-based genotyping. The genetic diversity and population structure were investigated using SNP markers from 380 lettuce accessions, which were maintained by the United States Department of Agriculture [3, 32]. The genotype data have been successfully used to identify lettuce cultivars, indicating that SNP markers can be useful for the rapid evaluation of genetic variation and population structure in the lettuce germplasm collection [18, 59]. In addition, research on genome-wide marker-trait association in lettuce has been conducted [32–35, 62, 63]. However, cultivar identification of commercial lettuce using SNP markers in Korea is still in its infancy.

Recent advancement of next-generation sequencing (NGS) technologies has enabled researchers to analyze and utilize genetic resources efficiently [36, 37, 60]. NGS technology has also accelerated high-throughput and genome-wide SNP genotyping [36–40]. GBS, one of the widely used NGS methods, is a high-throughput and cost-effective approach for discovering genome-wide SNPs. GBS has been used for the examination of genetic diversity in various plants, and SNP data from GBS have been applied for diverse genetic studies, cultivar identification, and marker-assisted breeding (MAB) [13–16, 33, 37–47, 61]. In lettuce, GBS has been successfully used to provide a large number of highly informative genome-wide SNPs [33].

Currently, diverse automated platforms for high-throughput analysis have enabled the analysis of large amounts of data within a short period [48, 49]. For example, Fluidigm dynamic arrays adopt an automated PCR and a nanofluidic integrated fluid circuit (IFC) [50]. SNP genotyping and the development of SNP markers for cultivar identification using the Fluidigm platform are being widely used for various plants [13–16, 51–54]. However, SNP markers for identifying different cultivars of commercial lettuce in Korea have not been sufficiently developed.

In this study, we developed core sets of genome-wide SNP markers to identify cultivars of lettuce using the GBS and SNP-genotyping approach. We validated these core sets of SNPs to develop molecular markers for high-throughput analysis using the Fluidigm platform. We also evaluated the utility of core SNPs using genetic differentiation analysis. These developed SNP markers will be useful for database construction and will facilitate cultivar identification, purity testing, and breeding of lettuce.

Results

Genome-wide SNP discovery in commercial lettuce cultivars

Using the GBS approach, a total of 549 123 132 reads were generated with an average of 6 034 320 reads per individual cultivar from the 90 Korean commercial lettuce cultivars analyzed. After barcode and adapter sequences were trimmed and low-quality reads were filtered, 443 005 356 clean reads were obtained (Table 1). About 86% of the

Table 1. Summary of GBS data for 90 lettuce cultivars

Class	No.
Total number of raw reads	549 123 132
Average number of raw reads per cultivar	6 034 320
Total length of raw reads	55 461 436 332
Total number of trimmed reads	443 005 356
Total number of mapped reads	380 064 388
Total SNP	276 462
Filtered SNP	17 877

reads were successfully mapped to the *L. sativa* cv. *Salinas* (v8) reference genome, with an average read depth of 19X [55].

A total of 276 462 SNPs were identified through genome-wide SNP identification. Among the SNPs identified from 90 cultivars used in this study, transition (A/G or C/T) and transversion (A/C, A/T, C/G or G/T) SNPs accounted for 62.8% and 37.2%, respectively, with a transitions-to-transversions ratio of 1.69. Both types of transition SNPs (C/T and A/G) were detected in similar numbers. Among transversion SNPs, the A/T type showed higher numbers than other types (Table S1).

Low-quality SNPs and with significant polymorphism among the 90 cultivars were filtered out. After filtering, 17 877 high-quality SNPs (minor allele frequency (MAF) > 0.05; missing data < 30%) were identified. The chromosomal distribution of these 17 877 SNP loci and genes in the lettuce genome is depicted in Fig. 1a. Generally, the SNPs were evenly distributed across the chromosomes.

Out of the 17 877 SNPs identified, 12 959 SNPs (72.5%) were located in intergenic regions and 4928 (27.6%) in genic regions, of which 3502 (19.6%) were located in exons and 1416 (7.9%) were in introns (Table 2). From the 3279 filtered SNPs derived from coding sequences, 2139 (65.2%) were found to be synonymous SNPs that do not alter the amino acid sequences of the polypeptide, whereas 1140 (34.8%) were discovered to be non-synonymous SNPs that cause changes to the amino acid sequences (Table 2). Of the non-coding sequence SNPs, 689 (3.9%) were derived from upstream and downstream regions of the genes.

Furthermore, transition (A/G or C/T) and transversion (A/C, A/T, C/G or G/T) SNPs accounted for 68.5% and 31.5%, respectively, with a transitions-to-transversions ratio of 2.18 (Table S1). Both types of transition SNPs (A/G and C/T) were detected in similar numbers. Among transversion SNPs, the A/C type showed higher numbers than other types.

Genetic diversity within lettuce cultivars

The level of genetic diversity, the polymorphic information content (PIC) values, MAF, and the heterozygosity of the 17 877 SNPs filtered from 90 lettuce cultivars were calculated (Table S2). As a result, PIC values ranged from 0.10 to 0.38, with an average PIC of 0.27. MAF for the selected SNP markers was from 0.06 to 0.50, with an average of 0.24, and the heterozygosity ranged from 0.10 to 0.38, with an average of 0.27. The selected 17 877 SNPs were used for the genetic analyses, including the phylogenetic analysis, principal component analysis (PCA), and population structure analysis.

A phylogenetic tree was constructed using the neighbor-joining method. The result showed that the 90 cultivars were classified into three divergent groups (Fig. 1b): 18 cultivars were classified into Cluster I (Red), 35 cultivars into Cluster II (Blue), and the 37 cultivars into Cluster III (Green). Generally, lettuce cultivars were clustered according to their horticultural types. Cluster I only comprised cultivars of the Cos (Romaine) type, and 27 out of 28 lettuce cultivars of Frisée d'Amerique type as well as several cultivars of the Cos type were clustered into Cluster II. In Cluster III, cultivars of the Cos, butterhead and Iceberg types were clustered, and there was a tendency to form subgroups by the horticultural types.

After conducting PCA using the 17 877 SNPs to investigate the genomic differences of lettuce cultivars, the 90 lettuce cultivars were classified into three groups (Fig. 1c). The top two principal components (PC1 and PC2) accounted for 19.9% of the genetic variation among the 90 cultivars. In addition, PC3 explained 4.4% of the observed variances (Data not shown). With a few exceptions, the phylogenetic relationships of different lettuce groups were in good agreement with the PCA results (Fig. 1b, c), resulting in three major clusters for the 90 cultivars, and each cluster generally comprised the same horticultural type. The majority of the cultivars in Cluster I was Cos type, while Cluster II contained 27 of the 28 Frisée d'Amerique type accessions. Cluster III included all six butterhead type accessions and two Iceberg type accessions, as well as several Cos type accessions. Seven accessions (LS003, LS019, LS020, LS025, LS053, LS054, and LS071) showed different grouping in phylogenetic (Cluster III) and PCA (Cluster I) analyses.

The population structure analysis for the 90 lettuce cultivars using the 17 877 filtered SNPs determined the optimal number of populations (K = 4) corresponding to the highest peak in the Delta-K graph (Fig. 1d and Fig. S1). The result suggested that genetic variations in the 90 cultivars can be divided into four major clusters, which was similar to the results of the PCA and phylogenetic analysis. Cluster I was composed of a mixture of 12 lettuce cultivars consisting of the Cos, Iceberg, multi-divided type, and unknown morphological type. Cluster II consisted of 17 lettuce cultivars, all of which were of the Cos type. All six Butterhead-type accessions, one Cos type accession, and one Lollo type accession were in Cluster III. All accessions of Frisée d'Amerique type were in Cluster IV. Most of the Cos-type accessions were found in two clusters, Cluster II (17 Cos-type) and Cluster IV (23 Cos-type), whereas four Cos-type accessions were found in Cluster I and one in Cluster III.

The second-highest peak in the Delta-K graph was found when K = 6, assuming six subgroups among the

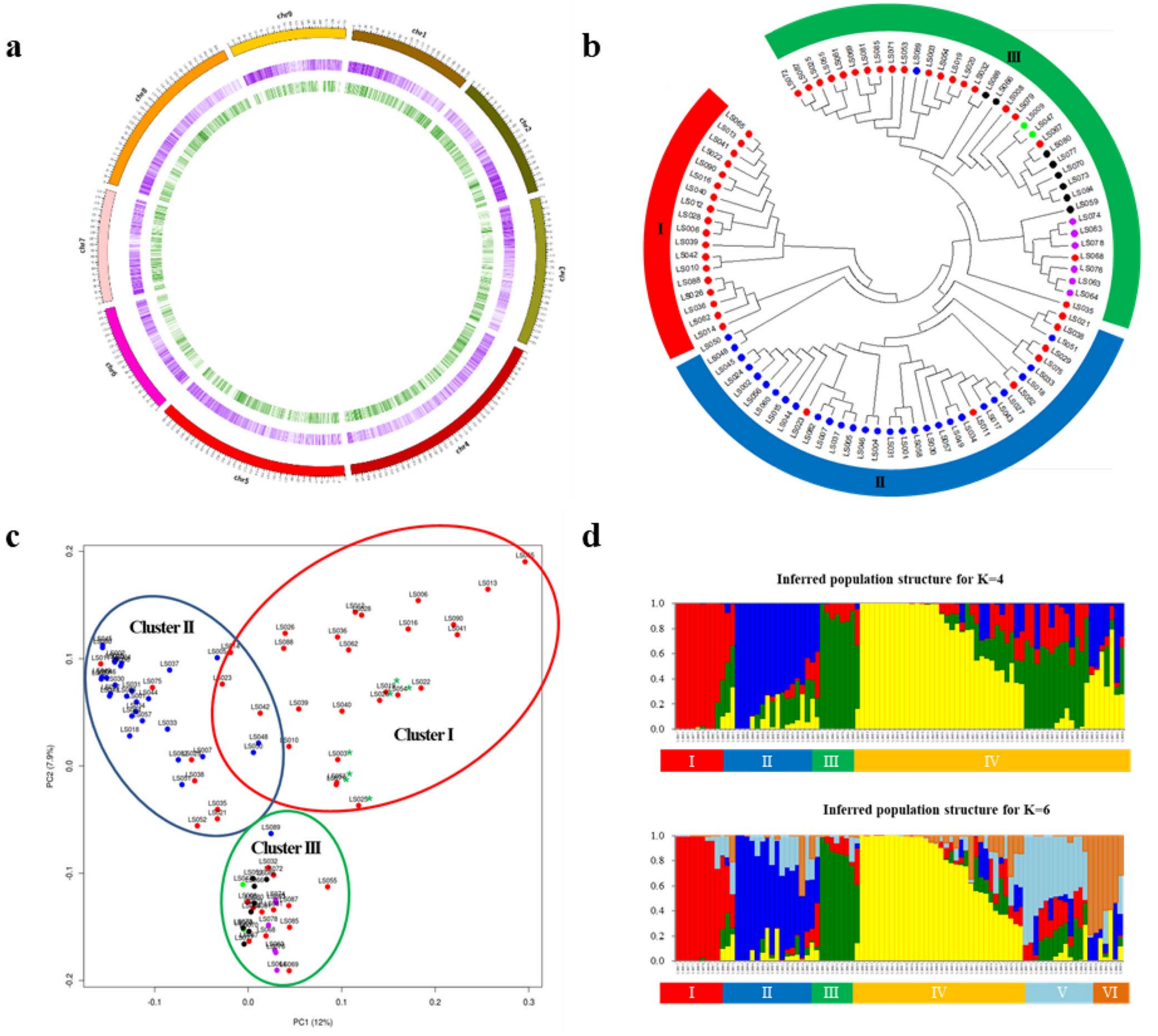


Figure 1. Chromosomal distribution and genetic diversity analyses of 17,877 SNP loci from 90 lettuce cultivars. a Distribution of genes and 17,877 SNP loci in lettuce genome. The middle purple circle illustrates gene distribution and the innermost green circle illustrates SNP distribution per 500kb. b Neighbor-joining phylogenetic tree using 17,877 SNPs in 90 lettuce cultivars. c Principal component analysis of the 90 lettuce cultivars genotyped with 17,877 SNPs. d Population structure of the 90 lettuce cultivars using 17,877 SNPs based on the STRUCTURE output for K=4 and K=6. The x axis shows the different lettuce cultivar and y axis represents co-ancestary coefficient. In b, c, colors of dots reflect the morphological types of lettuce; Cos (red), Iceberg (green), Frisée d'Amérique (blue), Butterhead (purple), Lollo, Multi-divided, and Oakleaf type (black). Differently colored outlines or bars indicate the main clusters.

90 cultivars (Fig. 1d and Fig. S1). No dramatic changes in clustering were observed when K was increased from 4 to 6, except for Cluster IV, with 29 Frisée d'Amerique type, 23 Cos-type, and one Oakleaf-type accessions being divided into three subgroups. Specifically, subgroup 1 of Cluster IV consisted of a majority of Frisée d'Amerique type lettuce accessions. Cos-type accessions were spread among all three subgroups: five in subgroup 1, 11 in subgroup 2, and seven in subgroup 3. Lettuce cultivars in subgroup 3 were composed of Cos-type accessions only.

Generally, cultivars with identical morphological types were grouped in the same cluster and a substantially close association was observed between SNP genotypes and horticultural types. In addition, the clusters obtained from the phylogenetic tree using the NJ method were in an agreement with those from STRUCTURE and PCA.

Development and validation of the core SNP assays for lettuce cultivar identification

To develop SNP markers that are effective for cultivar identification, we selected 294 SNP markers with a PIC value higher than 0.1 and with significant polymorphism between the 90 lettuce cultivars. Of these selected SNPs, 226 SNPs were in genic regions of the lettuce genome. To validate the selected SNPs, the sequence differences between cultivars were confirmed by Sanger sequencing

Table 2. The chromosomal location of 17 877 SNPs identified in 90 lettuce cultivars

Class	Total SNPs	Candidate SNPs	Core SNPs
Genic region	4928 (27.6%)	222 (75.5%)	159 (82.8%)
CDS - Synonymous	2139	133	96
CDS - Non-synonymous	1140	60	43
UTR	223	16	10
Intron	1416	13	10
Intergenic region	12 959 (72.5%)	72 (24.5%)	33 (17.2%)
Intergenic	12 270	67	31
Upstream gene	376	3	2
Downstream gene	291	2	0
Up/Downstream gene	22	0	0
Total number of SNPs	17 877	294	192

(Data not shown). Subsequently, primer sets for the Fluidigm assay were designed for each confirmed SNP. Fluidigm-based genotyping was conducted for the 90 lettuce cultivars used for GBS and the additional five commercial cultivars from the Netherlands. Genotype calls from the SNP assay for 95 lettuce cultivars are shown in the scatter plot (Fig. S2). Homozygous types, XX and YY, were labeled with fluorescent dyes, FAM or HEX, respectively, represented by red and green points. Heterozygous marker type (XY) was labeled with both fluorescent dyes FAM and HEX, represented by blue points. Among 294 SNP markers, SNPs which showed clear separation between two homozygous genotypes were selected for the development of the core marker sets (Fig S2a, b), and SNPs with unusual clustering patterns including heterozygous genotypes were filtered out (Fig. S2c). The automatically-called heterozygous genotypes by the software, as shown in Fig. S2b, were manually changed to homozygote genotypes.

The 294 filtered markers were effective in distinguishing 84 (88.4%) of the 95 lettuce cultivars. Besides, the samples which were not separated by the 294 markers could be used as a "reference variety" for each matching sample in DUS testing. To develop the core markers for cultivar identification via the Fluidigm system, sets of core markers were selected based on polymorphisms from the lettuce SNP data mined from GBS considering high polymorphism based on PIC value (Table S2). The average PIC value of the 192 core markers was 0.32, ranging from 0.1 to 0.38. The PIC values of the 96 core markers ranged from 0.11 to 0.38, with an average of 0.31. Similarly, the PIC value of the 48 core markers ranged from 0.11 to 0.38, with an average of 0.33. The average PIC value of the 24 core markers was 0.33, ranging from 0.23 to 0.37. These SNP sets are suitable for high-throughput systems such as the IFC platforms of Fluidigm genotyping assays. The selected SNP markers and their related gene annotations are summarized in Table S2.

To evaluate the performance for cultivar identification, a phylogenetic tree was constructed using the selected 192 SNP markers based on the genotyping results (Fig. 2a, Table S3). The 192 markers identified 84 (88.4%) of the 95 lettuce cultivars, like the 294 markers mentioned above, and were separated into three main clusters. The Frisée d'Amérique type was found almost exclusively in Cluster I, except for three cultivars that were grouped into other clusters. The cultivars of the Cos type were included in all clusters but were found mainly clustered in Cluster II. The Oakleaf types tended to be clustered with the Cos types, and the rest of the types tended to be clustered with each other.

Among the 192 SNP markers, the 96 core SNPs that showed significant polymorphism among the 95 cultivars were selected for further lettuce cultivar identification (Fig. 2b). Genetic differentiation using the 96 core SNPs was compared with those of distinct SNPs derived from GBS. As shown in Fig. 2b, the phylogenetic tree displayed three main clusters that are assigned different colors. The result showed that the accessions of identical horticultural types were included in the same cluster. All Frisée d'Amérique-type accessions were included in Cluster I, mainly in the subgroup Cluster I-1. Whereas, Cos-type accessions were spread into every subgroup: 17 in Cluster I, 24 in Cluster II, and four in Cluster III. Cluster I-4 comprised all accession types including Iceberg, Lollo, and multi-divided types, and six accessions of an unknown type. Cluster II included 19 Cos-type accessions and one Oakleaf type. Notably, Cluster II-2 contained all Butterhead-type accessions, and the accessions of an unknown type were clustered in Cluster I-4 and III. Representatively, the number of subgroups of 95 lettuce accessions was estimated using 96 core SNP markers (Fig. S3). The population structure analysis based on the genotyping results of 96 core SNPs indicated that the optimal number of subpopulations was three or seven, and this result was consistent with the phylogenetic tree constructed with 96 SNP markers.

From the 96 core markers, 48 and 24 markers were selected and phylogenetic analysis was performed using these markers. The 48 and 24 markers also identified 84 (88.4%) of the 95 lettuce cultivars, like the 294 markers (Fig. 2c, d). The 95 cultivars were separated into three main clusters, and associations between SNP genotypes and horticultural types were also detected (Fig. 3).

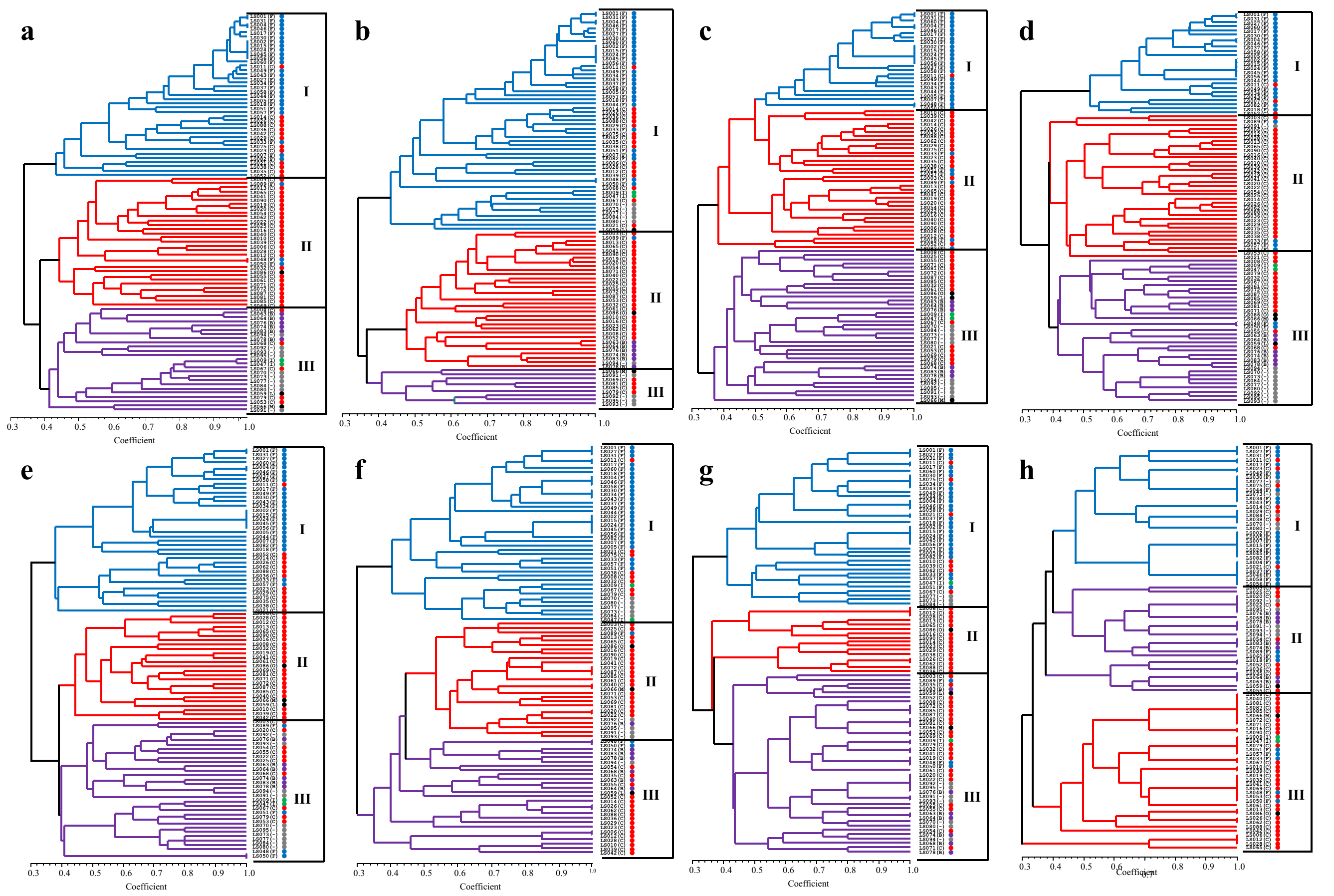


Figure 2. Phylogenetic trees of the 95 commercial lettuce accessions using the subsets of 192 (a), 96 (b), 48 (c), 24 (d), 18 (e), 12 (f), 9(g), and 6 (h) markers. The color of dot indicates horticultural type of each accession. The color of present Similarity coefficients are presented at the bottom of the trees.

Therefore, a set of 24 SNP markers would be sufficient for the selection of the reference varieties for the DUS examination, which would support UPOV-based PVP system. The genotyping results and the primer sequences of the 24 core SNP markers developed in this study have been provided in Table 3 and 4.

Additionally, a smaller number of markers were selected and analyzed to calculate the minimal number of SNP markers that can distinguish all tested cultivars. Eighteen markers were able to differentiate 84 (88.4%) of the 95 lettuce cultivars, similar to the 294 markers, while 12 markers were able to identify 67 (70.5%) of the 95 cultivars (Fig. 2e, f). Nine and six SNP markers detected genetic variations to distinguish 48 (50.5%) and 16 (16.8%) cultivars, respectively (Fig. 2 g, h).

To validate the developed 24 core markers, Fluidigm-based genotyping was performed for the sets of DNA mixtures and original samples and their genotyping results were compared. Fig. 4 shows the genotyping results of the original samples and their mixture. Original data showed two clusters corresponding to two homozygous genotypes (XX, red; YY, green) (Fig. 4a). However, heterozygous genotypes (XY, blue) were identified from some DNA mixtures (Fig. 4b). Among the reactions in which the genotypes of the two samples were different from each other, 93.3% genotypes of the mixture were heterozygous. Thus, we have determined that the developed markers could identify not only homozygous lines but also heterozygous lines.

Discussion

The main goal of this study was to develop a core set of SNP markers suitable for the Fluidigm assay to create a fast and high-throughput screening system for the identification of lettuce cultivars. In this study, the selected 17 877 SNPs using the GBS approach could successfully differentiate 90 commercial lettuce cultivars in Korea. A similar classification pattern was observed with phylogenetic analysis and PCA where both analyses classified the 90 lettuce cultivars into three distinct groups (Fig. 1). In both analyses, most of the cultivars included in Cluster I were Cos type, while Cluster II included the Frisée d'Amerique type as the majority. In Cluster III, Cos, Butterhead, and Iceberg types were present, and subgroups were formed within a horticultural type. Population structure analysis identified four genetic clusters, which nearly corresponded to the grouping of the accessions from phylogenetic and PCA analyses. Lettuce cultivars were generally classified into three main clusters according to their horticultural types. Therefore, these results demonstrate that the SNPs can be variable

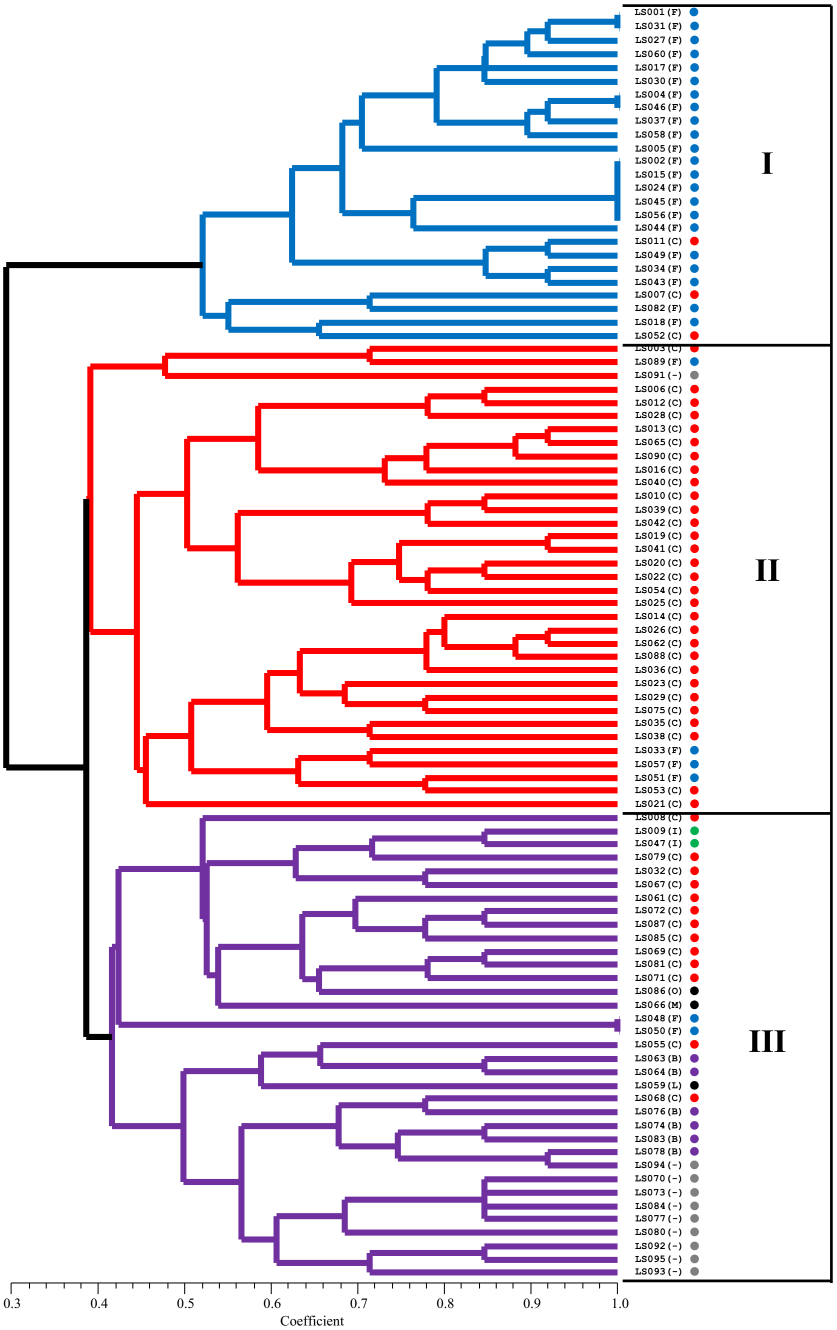


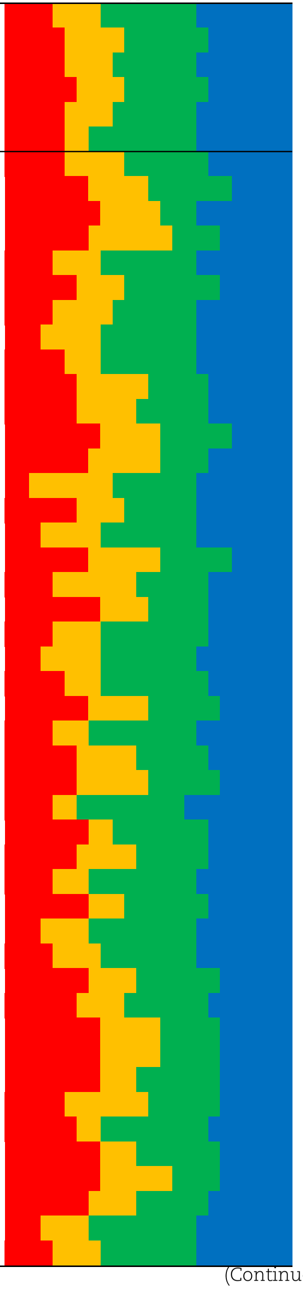
Figure 3. The phylogenetic tree of 95 commercial cultivars using 24 core SNP markers. The color of circle next to names of varieties indicate horticultural types; Cos type (red), Iceberg type (green), Frisée d'Amérique type (blue), Butterhead type (purple), unknown (gray), other types including Lollo, Multi-divided, and Oakleaf type (black). All three distinct groups were assigned by different colors. Jaccard's similarity coefficients are presented at the bottom of the trees.

resources to reveal the association between genomic variations and horticultural types of lettuce.

Core markers for Fluidigm SNP genotyping were selected based on the PIC values, which provide information on the extent of polymorphism revealed by the DNA marker and supports estimating relationships between cultivars [64, 65]. In general, the PIC value of multi-allelic markers, such as RAPD [56], AFLP [57, 58], and SSR markers can be as high as 0.5–1.0, while the PIC values of bi-allelic SNP markers range from 0–0.5 [66–68]. The comparatively high PIC value (mean PIC = 0.33) for the present core marker sets in this study will warrant that their classification accuracy. A subset of 294 SNPs selected from the 17 877 SNPs showed the genetic differentiation, as well as the distinctness and uniformity, of 84 (88.4%) out of 95 cultivars, including 90 Korean and five Dutch cultivars. All cultivars that were not distinguished by the SNP markers were Frisée d'Amerique type. Since they were mostly developed by the same companies, their genetic similarities are probably due to the similar breeding program including the use of the same inbred line.

Table 3. The genotypes of 24 core SNP markers in 95 commercial lettuce varieties

Horticutural type	Sample Code	SNP001	SNP002	SNP005	SNP010	SNP025	SNP038	SNP057	SNP063	SNP079	SNP082	SNP084	SNP120	SNP133	SNP135	SNP139	SNP140	SNP141	SNP147	SNP149	SNP158	SNP166	SNP177	SNP180	SNP181	Genotype Density[a]
Butterhead	LS063	AA	GG	TT	CC	CC	CC	TT	AA	GG	GG	CC	TT	GG	CC	GG	AA	GG	CC	CC	TT	AA	GG	GG	CC	
	LS064	AA	GG	TT	CC	CC	TT	TT	AA	AA	GG	CC	TT	GG	CC	GG	AA	GG	CC	CC	TT	AA	GG	GG	CC	
	LS074	GG	GG	TT	AA	CC	CC	TT	AA	GG	AA	TT	TT	GG	CC	GG	AA	GG	CC	CC	CC	AA	GG	GG	CC	
	LS076	AA	GG	TT	AA	TT	TT	TT	AA	GG	AA	CC	CC	GG	CC	GG	AA	GG	CC	CC	CC	AA	GG	GG	CC	
	LS078	GG	GG	CC	AA	TT	CC	TT	AA	GG	AA	TT	CC	GG	CC	GG	AA	GG	CC	CC	TT	AA	GG	GG	CC	
	LS083	GG	GG	TT	AA	CC	CC	TT	AA	GG	AA	CC	CC	GG	CC	GG	AA	GG	CC	CC	CC	AA	GG	GG	CC	
Cos	LS003	GG	CC	CC	AA	CC	TT	TT	GG	AA	AA	CC	TT	GG	CC	GG	AA	GG	TT	CC	TT	GG	AA	GG	CC	
	LS006	GG	CC	TT	AA	CC	TT	TT	GG	AA	AA	CC	TT	TT	AA	AA	GG	GG	CC	CC	CC	GG	AA	AA	CC	
	LS008	AA	GG	CC	AA	TT	TT	GG	AA	GG	GG	CC	TT	GG	AA	AA	GG	AA	TT	AA	CC	AA	GG	GG	TT	
	LS010	AA	GG	TT	AA	TT	TT	TT	AA	AA	AA	TT	CC	TT	AA	AA	GG	GG	TT	CC	CC	GG	GG	GG	CC	
	LS011	GG	CC	TT	CC	CC	CC	GG	GG	GG	AA	TT	CC	TT	AA	GG	AA	GG	TT	CC	CC	GG	AA	GG	CC	
	LS012	GG	CC	TT	AA	CC	TT	TT	GG	AA	AA	CC	CC	TT	AA	AA	GG	GG	CC	CC	CC	GG	AA	GG	CC	
	LS013	GG	CC	TT	AA	TT	TT	TT	GG	AA	AA	CC	TT	GG	CC	AA	GG	GG	CC	CC	CC	GG	GG	GG	CC	
	LS014	GG	CC	TT	CC	TT	CC	TT	GG	AA	AA	TT	CC	GG	CC	GG	AA	GG	TT	CC	CC	GG	GG	GG	CC	
	LS016	GG	CC	CC	AA	TT	TT	TT	GG	AA	AA	CC	CC	GG	AA	AA	GG	GG	CC	CC	CC	GG	GG	GG	CC	
	LS019	AA	GG	TT	AA	TT	TT	TT	AA	AA	AA	TT	TT	GG	CC	AA	GG	GG	CC	CC	CC	GG	GG	GG	CC	
	LS020	AA	GG	TT	AA	TT	TT	TT	AA	AA	AA	CC	TT	GG	CC	GG	AA	GG	CC	CC	CC	GG	GG	GG	CC	
	LS021	GG	CC	TT	AA	TT	CC	TT	AA	AA	AA	CC	CC	TT	AA	GG	AA	GG	TT	AA	CC	AA	GG	GG	CC	
	LS022	AA	GG	TT	AA	TT	TT	TT	AA	AA	AA	CC	TT	GG	CC	GG	AA	GG	TT	AA	CC	GG	GG	GG	CC	
	LS023	GG	CC	TT	CC	TT	TT	TT	GG	GG	AA	TT	CC	TT	CC	GG	AA	GG	TT	CC	CC	GG	GG	GG	CC	
	LS025	AA	GG	CC	AA	TT	TT	TT	GG	AA	AA	CC	TT	GG	CC	GG	AA	GG	CC	AA	CC	GG	GG	GG	CC	
	LS026	GG	CC	TT	CC	TT	CC	TT	GG	AA	AA	TT	CC	GG	CC	AA	GG	GG	TT	CC	CC	GG	GG	GG	CC	
	LS028	GG	CC	TT	AA	CC	TT	TT	GG	AA	AA	CC	TT	TT	AA	AA	GG	GG	CC	CC	CC	GG	AA	AA	TT	
	LS029	GG	CC	TT	CC	TT	TT	TT	AA	AA	AA	TT	CC	GG	CC	GG	AA	GG	TT	CC	TT	GG	GG	GG	CC	
	LS032	AA	GG	TT	AA	TT	TT	GG	AA	GG	AA	CC	TT	GG	CC	AA	GG	AA	CC	AA	CC	AA	GG	GG	CC	
	LS035	GG	CC	TT	CC	CC	CC	TT	AA	AA	AA	CC	TT	GG	CC	GG	AA	GG	TT	CC	CC	GG	GG	GG	CC	
	LS036	GG	CC	TT	CC	CC	TT	TT	GG	AA	AA	TT	CC	GG	CC	AA	GG	GG	TT	CC	CC	GG	GG	GG	CC	
	LS038	GG	CC	TT	CC	TT	CC	GG	AA	AA	AA	CC	CC	GG	CC	GG	AA	GG	TT	CC	CC	AA	GG	GG	CC	
	LS039	AA	GG	TT	AA	TT	TT	TT	AA	AA	AA	CC	CC	TT	AA	AA	GG	GG	CC	CC	CC	GG	GG	GG	CC	
	LS040	GG	CC	CC	AA	TT	TT	TT	AA	GG	AA	CC	CC	GG	CC	AA	GG	GG	CC	CC	CC	GG	GG	GG	CC	
	LS041	AA	GG	TT	AA	TT	TT	TT	AA	AA	AA	CC	TT	GG	CC	AA	GG	GG	CC	CC	CC	GG	GG	GG	CC	
	LS042	AA	GG	CC	CC	TT	TT	TT	AA	AA	AA	TT	CC	TT	AA	AA	GG	GG	TT	CC	CC	GG	GG	GG	CC	
	LS052	GG	CC	TT	CC	CC	CC	GG	GG	GG	GG	CC	CC	GG	CC	GG	AA	GG	CC	AA	CC	GG	AA	AA	TT	
	LS053	GG	CC	TT	AA	TT	CC	GG	AA	AA	AA	CC	CC	GG	CC	AA	GG	AA	CC	AA	CC	GG	GG	GG	CC	
	LS054	AA	GG	TT	AA	CC	TT	TT	AA	AA	AA	TT	TT	GG	CC	GG	AA	GG	CC	CC	CC	GG	GG	GG	CC	
	LS055	AA	GG	CC	CC	CC	TT	TT	AA	AA	GG	CC	TT	GG	CC	GG	AA	GG	CC	CC	CC	GG	GG	GG	CC	
	LS061	AA	GG	TT	AA	CC	TT	TT	AA	AA	GG	CC	CC	GG	CC	AA	GG	AA	CC	AA	CC	GG	GG	GG	CC	
	LS062	GG	CC	TT	CC	TT	CC	TT	GG	AA	AA	TT	CC	GG	CC	AA	GG	GG	CC	CC	CC	GG	GG	GG	CC	
	LS065	GG	CC	TT	AA	CC	TT	TT	GG	AA	AA	CC	TT	GG	CC	AA	GG	GG	CC	CC	CC	GG	GG	GG	CC	
	LS067	AA	GG	TT	CC	TT	CC	GG	AA	GG	AA	CC	TT	TT	CC	AA	GG	AA	CC	AA	CC	AA	GG	GG	CC	
	LS068	AA	GG	TT	AA	CC	TT	TT	AA	GG	AA	TT	CC	GG	CC	GG	AA	GG	CC	CC	CC	AA	GG	GG	CC	
	LS069	GG	GG	TT	AA	TT	CC	TT	AA	AA	AA	CC	TT	GG	CC	AA	GG	AA	CC	AA	TT	AA	GG	GG	CC	
	LS071	GG	GG	CC	AA	TT	TT	TT	AA	AA	AA	TT	TT	GG	CC	AA	GG	AA	CC	AA	CC	AA	GG	GG	CC	
	LS072	AA	GG	CC	AA	TT	TT	TT	AA	AA	GG	CC	CC	GG	CC	AA	GG	AA	CC	AA	CC	AA	GG	GG	CC	
	LS075	GG	CC	TT	CC	TT	CC	TT	AA	AA	AA	TT	CC	TT	AA	GG	AA	GG	TT	CC	TT	GG	GG	GG	CC	
	LS079	AA	GG	TT	CC	TT	CC	GG	AA	AA	GG	CC	CC	GG	CC	AA	GG	AA	CC	CC	CC	AA	GG	GG	CC	
	LS081	GG	GG	CC	AA	TT	CC	TT	AA	AA	AA	CC	TT	GG	CC	AA	GG	AA	CC	AA	CC	AA	GG	GG	CC	
	LS085	AA	GG	CC	AA	TT	TT	TT	AA	AA	GG	CC	TT	GG	CC	AA	GG	AA	TT	AA	TT	AA	GG	GG	CC	
	LS087	AA	GG	CC	AA	TT	TT	TT	AA	AA	GG	CC	TT	GG	CC	AA	GG	AA	CC	AA	CC	GG	GG	GG	CC	
	LS088	GG	CC	CC	CC	TT	CC	TT	GG	AA	AA	TT	CC	GG	CC	AA	GG	GG	TT	CC	CC	GG	GG	GG	CC	
	LS090	GG	CC	CC	AA	TT	TT	TT	GG	AA	AA	CC	TT	GG	CC	AA	GG	GG	CC	CC	CC	GG	GG	GG	CC	



(Continued)

Table 3. Continued

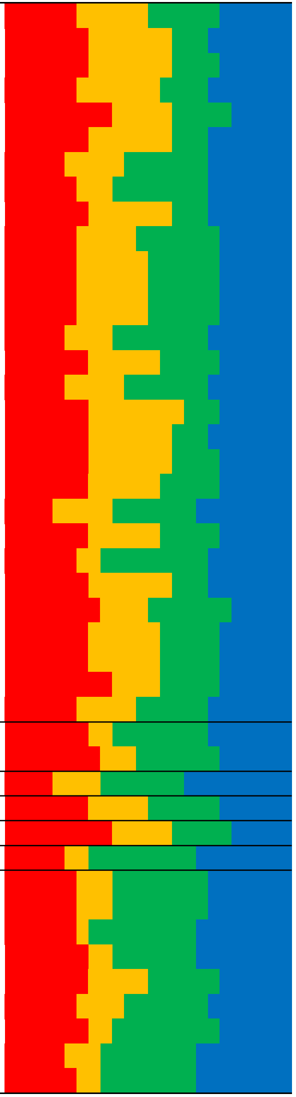
Horticutural type	Sample Code	SNP001	SNP002	SNP005	SNP010	SNP025	SNP038	SNP057	SNP063	SNP079	SNP082	SNP084	SNP120	SNP133	SNP135	SNP139	SNP140	SNP141	SNP147	SNP149	SNP158	SNP166	SNP177	SNP180	SNP181	Genotype Density[a]
Frisée d'Amérique	LS001	GG	CC	TT	CC	CC	CC	GG	AA	GG	AA	TT	CC	TT	AA	GG	AA	GG	TT	CC	TT	GG	AA	AA	TT	
	LS002	AA	GG	TT	AA	TT	CC	GG	GG	GG	AA	TT	CC	TT	AA	GG	AA	GG	TT	CC	TT	GG	AA	AA	TT	
	LS004	GG	CC	TT	AA	TT	CC	GG	AA	GG	AA	TT	CC	TT	AA	GG	AA	GG	TT	CC	TT	GG	AA	AA	TT	
	LS005	GG	CC	TT	AA	TT	TT	GG	GG	GG	AA	CC	CC	TT	AA	GG	AA	GG	TT	CC	TT	GG	AA	AA	TT	
	LS007	GG	CC	TT	AA	TT	CC	GG	GG	AA	AA	TT	CC	TT	AA	GG	AA	GG	CC	AA	CC	AA	AA	AA	TT	
	LS015	AA	GG	TT	AA	TT	CC	GG	GG	GG	AA	TT	CC	TT	AA	GG	AA	GG	TT	CC	TT	GG	AA	AA	TT	
	LS017	GG	CC	TT	CC	CC	CC	GG	GG	GG	AA	TT	CC	TT	AA	GG	AA	GG	TT	CC	CC	GG	AA	AA	TT	
	LS018	GG	CC	TT	CC	CC	CC	GG	GG	GG	AA	TT	CC	GG	AA	GG	AA	GG	CC	CC	CC	AA	AA	AA	TT	
	LS024	AA	GG	TT	AA	TT	CC	GG	GG	GG	AA	TT	CC	TT	AA	GG	AA	GG	TT	CC	TT	GG	AA	AA	TT	
	LS027	GG	CC	TT	CC	CC	CC	GG	AA	GG	AA	TT	CC	TT	AA	GG	AA	GG	TT	CC	CC	GG	AA	AA	TT	
	LS030	GG	CC	TT	CC	TT	CC	GG	AA	GG	AA	TT	CC	TT	AA	GG	AA	GG	TT	CC	CC	GG	AA	AA	TT	
	LS031	GG	CC	TT	CC	CC	CC	GG	AA	GG	AA	TT	CC	TT	AA	GG	AA	GG	TT	CC	TT	GG	AA	AA	TT	
	LS033	GG	CC	TT	CC	TT	CC	GG	AA	AA	AA	TT	CC	TT	AA	AA	GG	AA	TT	CC	TT	GG	GG	GG	CC	
	LS034	GG	CC	TT	CC	TT	CC	GG	AA	GG	AA	TT	CC	TT	AA	GG	AA	GG	CC	CC	CC	GG	AA	GG	CC	
	LS037	GG	CC	TT	AA	TT	CC	GG	AA	GG	AA	CC	CC	TT	AA	GG	AA	GG	TT	CC	TT	GG	AA	AA	TT	
	LS043	GG	CC	TT	CC	TT	CC	GG	AA	GG	AA	TT	CC	TT	AA	GG	AA	GG	TT	CC	CC	GG	AA	GG	CC	
	LS044	AA	GG	TT	CC	TT	TT	GG	AA	GG	AA	TT	CC	TT	AA	GG	AA	GG	TT	CC	TT	GG	AA	AA	TT	
	LS045	AA	GG	TT	AA	TT	CC	GG	GG	GG	AA	TT	CC	TT	AA	GG	AA	GG	TT	CC	TT	GG	AA	AA	TT	
	LS046	GG	CC	TT	AA	TT	CC	GG	AA	GG	AA	TT	CC	TT	AA	GG	AA	GG	TT	CC	TT	GG	AA	AA	TT	
	LS048	AA	GG	TT	AA	TT	CC	TT	AA	GG	AA	TT	CC	GG	CC	AA	GG	AA	CC	CC	TT	GG	GG	AA	TT	
	LS049	GG	CC	TT	CC	TT	CC	GG	GG	GG	AA	TT	CC	TT	AA	GG	AA	GG	TT	CC	CC	GG	AA	GG	CC	
	LS050	AA	GG	TT	AA	TT	CC	TT	AA	GG	AA	TT	CC	GG	CC	AA	GG	AA	CC	CC	TT	GG	GG	AA	TT	
	LS051	GG	CC	TT	CC	TT	CC	GG	AA	AA	AA	CC	CC	GG	AA	AA	GG	AA	CC	CC	CC	GG	GG	GG	CC	
	LS056	AA	GG	TT	AA	TT	CC	GG	GG	GG	AA	TT	CC	TT	AA	GG	AA	GG	TT	CC	TT	GG	AA	AA	TT	
	LS057	GG	CC	TT	CC	TT	CC	GG	AA	AA	AA	TT	CC	GG	AA	AA	GG	AA	TT	CC	CC	AA	AA	GG	CC	
	LS058	GG	CC	TT	AA	TT	CC	GG	AA	GG	AA	TT	CC	TT	AA	GG	AA	GG	CC	CC	TT	GG	AA	AA	TT	
	LS060	GG	CC	TT	AA	CC	CC	GG	AA	GG	AA	TT	CC	TT	AA	GG	AA	GG	TT	CC	TT	GG	AA	AA	TT	
	LS082	AA	GG	TT	AA	TT	CC	GG	GG	GG	AA	CC	CC	TT	AA	GG	AA	GG	CC	AA	CC	AA	AA	AA	TT	
	LS089	AA	GG	TT	AA	CC	TT	TT	GG	AA	AA	CC	CC	GG	CC	GG	AA	GG	TT	CC	TT	GG	AA	GG	CC	
Iceberg	LS009	AA	GG	TT	CC	TT	CC	GG	AA	GG	GG	CC	CC	GG	CC	AA	GG	AA	CC	AA	CC	AA	AA	GG	CC	
	LS047	AA	GG	TT	CC	TT	CC	TT	AA	GG	GG	CC	CC	GG	AA	AA	GG	AA	CC	AA	CC	AA	AA	GG	CC	
Lollo	LS059	GG	GG	TT	CC	CC	TT	TT	AA	GG	GG	CC	CC	GG	CC	AA	GG	GG	CC	AA	TT	AA	GG	GG	CC	
Multi-divided	LS066	GG	CC	CC	AA	TT	TT	TT	AA	GG	AA	CC	CC	GG	CC	AA	GG	AA	CC	AA	TT	AA	GG	GG	TT	
Oakleaf	LS086	AA	GG	TT	AA	TT	TT	TT	GG	AA	AA	CC	TT	GG	CC	AA	GG	AA	CC	AA	CC	AA	GG	AA	CC	
Unknown	LS070	AA	GG	TT	CC	TT	CC	GG	AA	GG	AA	CC	CC	GG	CC	GG	AA	GG	CC	CC	CC	AA	GG	GG	CC	
	LS073	AA	GG	TT	CC	TT	CC	GG	AA	GG	AA	CC	CC	TT	AA	GG	AA	GG	CC	CC	CC	AA	GG	GG	CC	
	LS077	AA	GG	TT	CC	TT	CC	GG	AA	GG	AA	CC	CC	TT	CC	GG	AA	GG	CC	AA	CC	AA	GG	GG	CC	
	LS080	AA	GG	CC	CC	TT	CC	GG	AA	GG	GG	CC	CC	GG	CC	GG	AA	GG	CC	AA	CC	AA	AA	GG	CC	
	LS084	AA	GG	TT	CC	TT	CC	GG	AA	GG	AA	CC	CC	GG	AA	GG	AA	GG	CC	AA	CC	AA	GG	GG	CC	
	LS091	AA	GG	CC	AA	TT	CC	TT	AA	GG	AA	CC	CC	GG	CC	GG	AA	GG	TT	CC	TT	AA	AA	GG	TT	
	LS092	AA	GG	TT	AA	TT	CC	TT	AA	AA	GG	CC	TT	GG	CC	GG	AA	GG	CC	CC	CC	AA	GG	GG	CC	
	LS093	AA	GG	CC	AA	TT	CC	TT	AA	AA	AA	CC	CC	GG	CC	GG	AA	GG	CC	CC	CC	AA	GG	GG	CC	
	LS094	GG	GG	CC	AA	CC	CC	TT	AA	GG	AA	TT	CC	GG	CC	GG	AA	GG	CC	CC	TT	AA	GG	GG	CC	
	LS095	AA	GG	TT	AA	TT	CC	GG	AA	AA	GG	CC	CC	GG	CC	GG	AA	GG	CC	CC	CC	AA	GG	GG	CC	

Table 4. Summary of the information of 24 core SNP markers developed in this study

No.	Chro-mosome	Position	Allele	SNP location	giNumber	MAF[a]	PIC[b]	ASP1 sequence	ASP2 sequence	LSP sequence	STA sequence
SNP001	chr1	11 945 871	A/G	CDS	gi|357 464 531	0.49	0.37	CTCTTCGCCAGTGCCACT	TCTTCGCCAGTGCCACC	GTCAATGACGCCGGCGTT	CTTTGTCTGGGAGCTGCAC
SNP002	chr1	13 440 345	G/C	CDS	gi|1 137 180 833	0.48	0.37	AGGAGTATTAACAGCAGCTGTCC	AGGAGTATTAACAGCAGCTGTCG	CGGAACGTTGCCACTCGC	CCGATTGGAAACGACGGC
SNP005	chr1	44 033 573	C/T	Intron	gi|976 911 262	0.22	0.28	TGGCAGGTAACTAACTAACTGACTG	GTGGCAGGTAACTAACTAACTGACTA	ACGCCAATCCCCTGTATCACA	GCTTCTTTGGTGTCAATGTTGGT
SNP010	chr1	69 220 682	C/A	CDS	gi|976 926 764	0.26	0.31	CCTCCCATGAACAACAATCAAAACTAG	CCTCCCATGAACAACAATCAAAACTAT	GCTGCTCATGGCTCTCAAGTG	TGCTGCATGCAACCCC
SNP025	chr2	91 614 666	T/C	CDS	gi|976 924 026	0.22	0.28	CTTCTGATGCATGAAGACCATGT	CTTCTGATGCATGAAGACCATGC	GCAATCAGAAACTCTGAAAATGATGATTGGA	GTTGTTGTAAAACTTGCACTGCT
SNP038	chr2	117 891 323	C/T	Intergenic	-	0.44	0.37	ACCAAGGAACTAGAGGACGAAATC	AACCAAGGAACTAGAGGACGAAATT	CTGCATTCTCTTTCATTGCTTCCCT	AACTAATTGCGGGTCTTAATGCTAA
SNP057	chr3	68 819 733	G/T	CDS	gi|976 908 992	0.42	0.37	CATCATATACAGGTAGGGTGCACC	CATCATATACAGGTAGGGTGCACA	CCTGTACTTGCAGCCCGGT	CTTTCTCTGATTTGATAATTGTTTGATGTTCA
SNP063	chr3	144 267 704	A/G	Intergenic	-	0.21	0.28	GTCCACAACGTTTTTGGGGT	TCCACAACGTTTTTGGGGC	CCCTAATAGCCTACATGTATATAAAGGACCC	ACGTGTGAAAACCCTAAAGAACAT
SNP079	chr4	226 988 250	G/A	CDS	gi|1 098 831 891	0.49	0.37	CTACGAAAATAGAAAGATCAGAAGAAGCG	CCTACGAAAATAGAAAGATCAGAAGAAGCA	TGGGCTACATTTAGCACGGGA	CGTCAGTTACATCTGCAGCTA
SNP082	chr5	8 785 574	A/G	Intergenic	-	0.16	0.23	CACCCGCTACCCGATCA	CACCCGCTACCCGATCG	GGCCGCACTTATGGGAAGTT	CCTTGCGAAACTTGTGGCA
SNP084	chr5	36 437 007	C/T	Intron	gi|976 912 394	0.43	0.37	TGCCATCAGCTACAGTGAGTAATC	TGCCATCAGCTACAGTGAGTAATT	CTCTTGACTATCCTTTCAATTCATTAATACGAAACAG	CGTCTGACTTTAGCTGCAATGATTA
SNP120	chr5	284 742 114	C/T	CDS	gi|976 892 502	0.32	0.34	ATCAGCGGAACCCCCG	CATCAGCGGAACCCCCA	CCAGAGGGGATTGCTGCAT	TCCATAGGTCCTCACTTGTGC
SNP133	chr7	112 702 864	T/G	CDS	gi|976 916 844	0.39	0.36	CTCACCCTTTGTTTGAGCCAT	CTCACCCTTTGTTTGAGCCAG	GCAGCTTGTCCAAAGAGCCA	CTGCTCGGTTCTTTGATACCC
SNP135	chr7	133 109 575	C/A	UTR	gi|1 130 837 053	0.42	0.37	TTCTCAGCCTCGATTCCGG	GTTCTCAGCCTCGATTCCGT	GCTTCTTCTCCATCTCCGCCA	GGCGGAGATGACCCGT
SNP139	chr8	32 465 725	G/A	CDS	gi|976 920 382	0.45	0.37	AATGAAACGATTTGATTAGCTTCTCAAGTATC	AATGAAACGATTTGATTAGCTTCTCAAGTATT	TGGCAAGTCTATGAAGCAGCCT	TGCAGTTCACGAGGAGGT
SNP140	chr8	32 465 762	A/G	CDS	gi|976 920 382	0.45	0.37	CTGCAGTTCACGAGGAGGTAT	CTGCAGTTCACGAGGAGGTAC	GGGATACTTGAGAAGCTAATCAAATCGTTTC	TGAAGGACATATTTAGGATCTTTGTGC
SNP141	chr8	32 467 169	G/A	CDS	gi|976 920 382	0.18	0.26	CTCAATCAGTGATTGGTTCTCGATTTC	CTCAATCAGTGATTGGTTCTCGATTTT	GGCTTCACAGAGGCTTCTAATTTTGATG	GCTTCTCAGACTCTTTGCTTCC
SNP147	chr8	76 000 585	C/T	CDS	gi|976 929 628	0.4	0.37	ACAGAAGGAGGACGAAACAACG	ACAGAAGGAGGACGAAACAACA	TCGCAGCTTCGGATTCCTCT	GTCGAAGAAGAAGAAGAACAAACAGA
SNP148	chr8	78 404 531	C/T	CDS	gi|731 395 356	0.32	0.34	AACGTCATGTCCAGACGACTC	AAACGTCATGTCCAGACGACTT	CAGCAAATCGGTTGGTTGCAG	CAGCTTTTGTTGCTCCAATAACAA
SNP158	chr8	208 004 895	C/T	CDS	gi|976 919 073	0.25	0.31	TTTCATGCTTTTTAGGCTTCCCC	GTTTCATGCTTTTTAGGCTTCCCT	TGCCGGATCACCGAAGCTTA	CAGCCATGGATATGGATCCATT
SNP166	chr9	32 125 232	A/G	CDS	gi|976 901 682	0.36	0.36	GGCTGCTAGATTTCATTCTCACTGT	GCTGCTAGATTTCATTCTCACTGC	GGATGGTAGTACAACCGGGCA	GGGAAATACGTTGAGATACTGGATT
SNP177	chr9	77 623 861	G/A	CDS	gi|976 926 002	0.36	0.36	GAGTCAACTGTTTCCTGGTGTTC	TGAGTCAACTGTTTCCTGGTGTTT	CCGGTGCCACGTGTCC	TGATTAGCGCATTCACGTTTGT
SNP180	chr9	133 535 460	G/A	CDS	gi|976 909 018	0.26	0.31	AGCTGAAGCAGCGGTGG	AGCTGAAGCAGCGGTGA	GCCGCGCTCCGCCTA	AACGATAAATCGGGAGTGATAGAGA
SNP181	chr9	135 041 719	C/T	CDS	gi|976 927 416	0.26	0.31	AGGTGATCAAGAAGCTGGATGAC	AAGGTGATCAAGAAGCTGGATGAT	CCCAGGGTCAGGCTTTCG	ACTTAACCAAAGAAAAGAAGGTGATCA

[a] minor allele frequency
[b] polymorphic information content

If a minimal amount of markers can be used for cultivar identification without decreasing resolution, the cost, time, and labor required for cultivar identification will be significantly decreased, thus increasing the overall efficiency of the method. The core sets of markers (192, 96, 48, and 24 SNPs) selected from 294 filtered SNPs, and the subsets of core markers also successfully distinguished lettuce cultivars by identifying 84 (88.4%) of the 95 studied cultivars similar to the 294 markers (Fig. 2). Therefore, the core set of 24 markers developed in this study will be a useful tool for new cultivar registration and protection by supporting the selection of the "reference varieties" for DUS testing. The additional reduced number of marker sets were analyzed to identify the minimal number of SNP markers that can distinguish the tested cultivars. Eighteen markers also identified 88.4% of the studied cultivars and they can be substituted when needed to reduce the number of markers while maintaining the identification rates of the core marker set. The additional reduced number of marker sets such as 12, nine, and six SNP markers respectively distinguished 66 (69.5%), 47 (49.5%) and 15 (15.8%) cultivars, with lower identification rates proportional to the applied marker numbers in inverse order. Therefore, the subsets of SNP markers would also be useful for quick identification of lettuce cultivars. The application of the reduced number of SNP markers will enhance the efficiency of cultivar identification at a lower cost and with reduced effort.

In this study, the grouping results based on Fluidigm genotyping data showed a similar pattern to those obtained from 17 877 SNP markers from GBS data. The 24 core SNPs classified 95 lettuce cultivars into three distinct clusters and each cluster showed a tendency to be grouped with the same horticultural type (Fig. 3). This result showed that the developed SNP markers may be effectively used for genetic diversity and marker-trait association studies of lettuce. Although there is no direct evidence that the developed markers are associated with morphological traits, further in-depth studies on associations between genetic markers and phenotypic traits would allow better identification of cultivars. A previous study reported that 384 SNPs from 298 homozygous lettuce lines were used to assess the association between SNPs and ten horticultural traits, resulting in the detection of nine significant marker-trait associations [32]. Therefore, associations between molecular markers and morphological traits will contribute to the precise cultivar identification as a supplement to morphological analysis.

Since lettuce is a principally self-pollinated crop, homozygous genotypes are predicted to be predominant. Here, at first, the results obtained from automated calling data showed an unusually high portion of heterozygous SNPs. Therefore, all 192 SNPs were reanalyzed and the automatically-called heterozygous genotypes by the software were changed to homozygote genotypes. However, heterozygous genotypes at the polymorphic loci among 192 SNPs were still observed at a low level, accounting for 0.25% of the total genotyping results. The appearance of heterozygous genotypes in lettuce cultivars could have been caused by technical limitations, incomplete fixation of cultivars, seed purity problems. Therefore, the core 24 SNPs, which showed a clear separation between two homozygous genotypes were selected for lettuce cultivar identification (Table 3,4).

To validate that the 24 core markers accurately differentiated between homozygous and heterozygous genotypes, additional genotyping was performed (Fig. 4). As F_1 cultivars were not available, we generated artificial heterozygous lines by mixing equal amounts of DNAs from two homozygous lines with different alleles. Of the genotyping results, 93.3% were heterozygous. This demonstrated that the developed markers could identify not only homozygous lines but also heterozygous lines.

DUS testing is required for plant cultivar registration and protection, and SNP markers have successfully supported the management of reference collections, including the selection of a "similar variety" as the test cultivar with the highest genetic similarity [7]. As mentioned above, the 24 core markers developed in this study and the total of 294 filtered SNPs that identified 89.5% of the 95 lettuce cultivars can be used to reduce higher numbers of cultivars and/or breeding lines candidates for DUS. For instance, genetically similar cultivars and/or breeding lines selected for further DUS testing can be narrowed down for morphological characterization saving resources and time to seed companies. The marker sets developed in this study can be used to select the reference varieties for use in the DUS test by screening genetically similar cultivars. Therefore, the developed markers will be valuable resources to improve the DUS system and construct a DNA-based PVP system.

In conclusion, genome-wide SNP marker discovery via GBS and SNP genotyping using the Fluidigm system was successfully applied to assess the genetic diversity of lettuce (*L. sativa*) and validate the selected core sets of markers for cultivar identification as a part of the DUS test of lettuce. Based on these results, we constructed SNP database for lettuce cultivar identification using the genotyping results of Korean commercial lettuce cultivars. The constructed SNP database will support cultivar identification, population structure analysis, lettuce breeding, and DUS test for plant cultivar protection and enforcement of the right of breeders. Further research of SNP genotyping using the core marker sets developed in this study for additional lettuce cultivars will facilitate the utility of the markers to identify diverse cultivars worldwide. Additionally, a genome-wide association study and quantitative trait locus mapping are needed to understand the association between marker and trait.

Materials and methods

Plant materials and DNA extraction

Ninety Korean commercial lettuce cultivars were used to obtain whole-genome data of lettuce cultivars, and

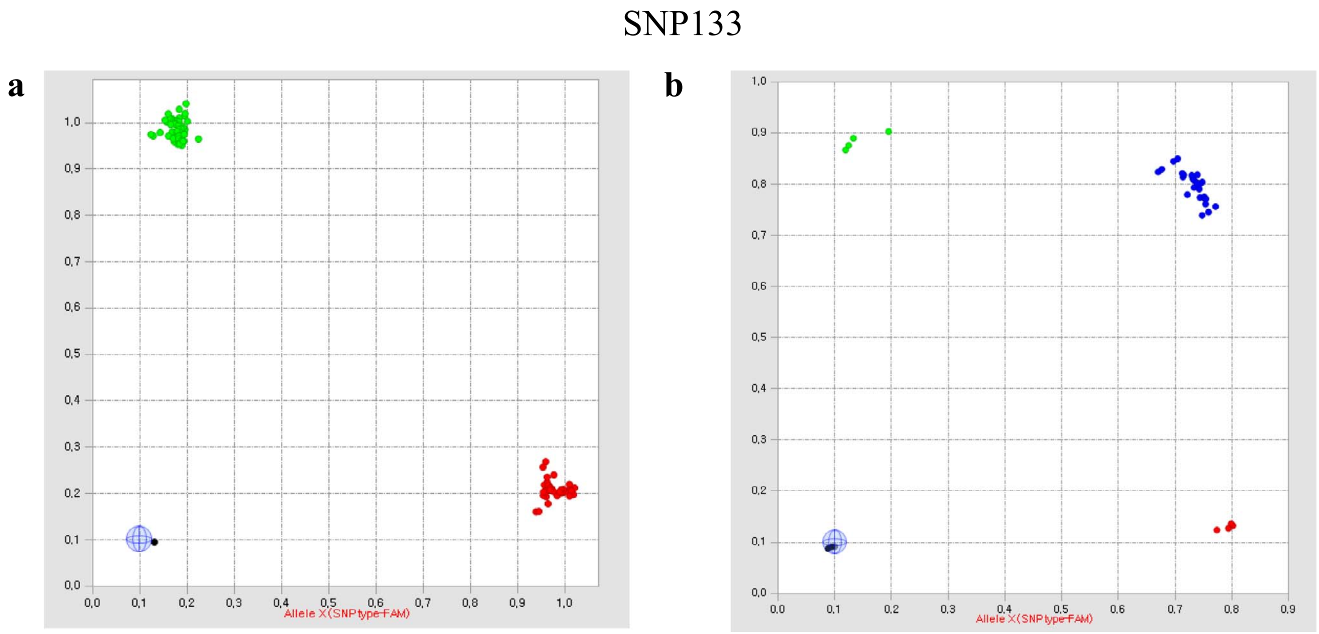


Figure 4. Representative clustering patterns of homozygotes and the artificial heterozygotes. Colored dots presents the genotypes from the same SNP marker of the 95 lettuce cultivars (a) and DNA mixtures from two cultivars with different alleles (b). Each color code in the plots presents one of three genotypes: homozygote of allele 1 (red), homozygote of allele 2 (green), and heterozygote (blue).

five lettuce cultivars from the Netherlands were added for marker validation (Table S5). Their horticultural types were determined by the KSVS, Gimcheon, Korea, based on UPOV TG/13/10 guidelines for lettuce [4]. In this study, the Cos type was the most abundant, accounting for more than 50% of 90 cultivars, followed by the Frisée d'Amérique type and the Butterhead type. Seeds were obtained from each cultivar grown at a greenhouse in KSVS (Gimcheon, Korea), and young leaf tissue was collected from each of the 95 lettuce cultivars. Genomic DNA was extracted using a DNeasy 96 Plant kit (Qiagen, cat. no. 69181, Germany) according to the manufacturer's instructions. DNA concentration and quality were determined using NanoDrop 8000 (Thermo Scientific, USA). The extracted DNA was normalized and used for GBS library construction and SNP genotyping. For SNP genotyping using the Fluidigm system, the concentration of DNA samples was adjusted to 10 ng/μL.

GBS

GBS libraries were constructed for the 90 lettuce cultivars following the protocols described by Elshire et al. [69]. In brief, the genomic DNA of each cultivar was digested with *Ape*KI (New England Biolabs, Ipswitch, MA, USA) and ligated to a specific barcode adapter. Digested DNAs were pooled and purified with the QIAquick PCR Purification Kit (Qiagen, Valencia, CA, USA).

GBS libraries were sequenced using the HiSeq2000 (Illumina, Inc., San Diego, CA, USA) by SEEDERS Inc. (Daejeon, Korea) and demultiplexing was conducted based on barcode sequence information. Adapter trimming was performed using the Cutadapt (v.1.8.3) program [70] and sequence quality trimming was conducted using the DynamicTrim and LengthSort program of the SolexaQA (v.1.13) package [71]. The quality control standards were: i) minimum phred score of 20 and ii) minimum read length of 25. The cleaned reads were aligned to the reference genome sequence of lettuce [2] using Burrows-Wheeler Aligner (BWA 0.6.1-r104) [72] with default parameters except the following options: i) seed length (−l) = 30, ii) maximum differences in the seed (−k) = 1, iii) number of threads (−t) = 32, iv) mismatch penalty (−M) = 6, v) gap open penalty (-O) = 15, and vi) gap extension penalty (−E) = 8.

SNP calling

SNP calling was conducted by SEEDERS Inc. (Daejeon, Korea) with an in-house script [73] and SAMtools (v.0.1.16) [74] with default parameters, except the following options: i) minimum mapping quality for SNPs ≥30, ii) mapping quality for gaps ≥15, iii) read depth ≥ 3 and ≤ 190, iv) minimum InDel score for nearby SNP filtering ≥30. SNP matrix was generated and filtered with the following conditions: i) minimum depth ≥ 3, ii) MAF > 5%, and iii) missing data <30%, using in-house script [73]. The raw SNP positions identified from each sample were integrated, and the non-SNP loci were filled with the consensus sequence of the sample and the miscalled lloci were filtered out. The physical positions of the SNPs in the genome, such as coding sequence (CDS), intronic, or intergenic region, were identified.

Genetic differentiation analysis

To estimate genetic differentiation and the number of subpopulations among lettuce cultivars using the filtered

SNPs, the 95 lettuce accessions were analyzed and clustered via hierarchical clustering, PCA, and population structure analysis. Phylogenetic analysis was conducted using the NJ method implemented in Molecular Evolutionary Genetics Analysis 6 (MEGA 6) [75], and a phylogenetic tree was visualized using the bootstrap method. PCA was conducted based on 17 877 SNPs using the R package SNPRelate [76]. The population structure analysis was conducted using STRUCTURE software [77], and each number of assumed clusters (*K*) was set from 1 to 10. The optimal *K* value was calculated using the Delta-K method (Δ*K*) described by Evanno et al. [78].

To develop markers suitable for lettuce cultivar identification, we selected SNPs based on the PIC value and chromosomal position. The PIC value for SNP markers was calculated according to the following formula:

$$\mathrm{PIC} = 1 - \sum_{i=1}^{n} p_i^2 - \sum_{i=1}^{\mathbf{n}-1} \sum_{j=i+1}^{n} 2p_i^2 p_j^2$$

where n is the number of alleles and p_i and p_j are the frequency of the i th and j th allele, respectively.

SNP validation and genotyping

To validate SNPs discovered from GBS data, Sanger sequencing was conducted using the flanking sequences of the selected SNPs. For high-throughput analysis using the Fluidigm system, we converted SNPs from GBS into SNP type assays to be used in Fluidigm 192.24, 96.96, 48.48, or 24.192 dynamic arrays, which yielded data points with 192, 96, 48, or 24 markers, respectively. The primer sequences of the selected SNP markers are listed in Table S4. The selected SNPs were validated via a high-throughput Fluidigm Juno™ system (Fluidigm Corporation, San Francisco, CA, United States) with 95 lettuce cultivars, according to the manufacturer's instructions. To test the efficiency of the prepared SNP assays, DNAs of five cultivars from the Netherlands were also included in the Fluidigm assay (Table S5). Sets of specific target amplification (STA) primer, a locus-specific primer (LSP), and an allele-specific primer (ASP) were designed for each SNP using the Fluidigm D3 Assay Design (https://d3.fluidigm.com/; Fluidigm, South San Francisco, CA, USA) [50]. Template sequences were prepared at a length of 200–300 bp, including 100 bp upstream and downstream of the targeted SNP. The 96.96 IFC (Fluidigm, South San Francisco, CA, USA) was used to perform the SNP type assays according to the manufacturer's instructions [50]. The pre-amplification step was performed using LSP and STA primers and the pre-amplified products were amplified using a set of ASPs.

Further, fluorescence intensity was quantified using the Fluidigm EP1 (Fluidigm Corporation, San Francisco, CA, USA), and the SNPs were called using Fluidigm SNP Genotyping Analysis software v4.5.1 according to the manufacturer's protocol. The allelic data matrix of "1" or "0" was used for the population genetic analysis for the genotyping results. A phylogenetic tree was generated based on SNP markers via NTSYS-pc 2.2 program (Applied Biostatistic, New York, USA) [79] using sequential agglomerative hierarchical nested clustering analysis.

To demonstrate if the developed markers worked accurately, we conducted SNP genotyping with 24 core markers, as well as 20 sample sets with two cultivars and their DNA mixture, using the Fluidigm system. The genotype results of the DNA mixtures were compared with those of the original cultivars.

Acknowledgements

This work was supported by the Korea Institute of Planning and Evaluation for Technology in Food, Agriculture, Forestry and Fisheries (IPET) through the Agri-Bio industry Technology Development Program, funded by the Ministry of Agriculture, Food and Rural Affairs (MAFRA) (317011-4).

Author contributions

J.P. performed experiments and wrote the manuscript. M.K., E.S., K.S., W.L. designed and performed the experiments. K.K. analyzed the data and revised the structure of the manuscript. J.O., S.S., S.J., Y.P., G.L. critically revised the paper. M.K. performed data visualization. J.J. supervised the project. All authors reviewed, edited, and approved the final manuscript.

Data availability

GBS reads of the 90 Korean lettuce cultivars have been deposited in the National Center for Biotechnology Information (NCBI) with BioProject accession number PRJNA746621 (Release data: 01-01-2022). The genotypic data of the developed SNP markers for the 90 and 5 Korean and Dutch cultivars are included as supplementary information. All relevant data of this study are included in this published article and its supplementary information files.

Conflict of interest

The authors have declared that no competing interests exist.

Supplementary data

Supplementary data is available at *Horticulture Research* online.

References

1. Korean Statistical Information Service (KOSIS). http://kosis.kr/ (2020).
2. Reyes-Chin-Wo S, Wang Z, Yang X et al. Genome assembly with in vitro proximity ligation data and whole-genome triplication in lettuce. *Nat Commun*. 2017;**8**:14953.

3. Kwon S, Truco M, Hu J. LS germ OPA, a custom OPA of 384 EST-derived SNPs for high-throughput lettuce (*Lactuca sativa* L.) germplasm fingerprinting. *Mol Breed*. 2012;**29**:887–901.
4. UPOV, International Union for the Protection of New Varieties of Plants. *Lettuce Guidelines for the Conduct of Tests for Distinctness Uniformity and Stability*. UPOV/TG/13/11, https://www.upov.int/edocs/tgpdocs/en/tg013.pdf (2017).
5. Jamali SH, Cockram J, Hickey LT. Insights into deployment of DNA markers in plant variety protection and registration. *Theor Appl Genet*. 2019;**132**:1911–29.
6. UPOV, International Union for the Protection of New Varieties of Plants. *Guidance on the Use of Biochemical and Molecular Markers in the Examination of Distinctness, Uniformity and Stability (DUS)*. *TGP/15*, https://www.upov.int/edocs/tgpdocs/en/tgp_15.pdf (2013).
7. UPOV, International Union for the Protection of New Varieties of Plants. *Guidelines for DNA –Profiling: Molecular Marker Selection and Database Construction (BMT Guidelines)*. UPOV/INF/17/1, http://www.upov.int/edocs/infdocs/en/upov_inf_17_1.pdf (2010).
8. Reid A, Hof L, Felix G et al. Construction of an integrated microsatellite and key morphological characteristic database of potato varieties on the EU common catalogue. *Euphytica*. 2011;**182**:239–49.
9. Vélez MD, Ibáñez J. Assessment of the uniformity and stability of grapevine cultivars using a set of microsatellite markers. *Euphytica*. 2012;**186**:419–32.
10. Zhou H, Zhang P, Luo J et al. The establishment of a DNA fingerprinting database for 73 varieties of *Lactuca sativa capitate* L. using SSR molecular markers. *Hortic Environ Biotechnol*. 2019;**60**: 95–103.
11. Zhang S, Li B, Chen Y et al. Molecular-assisted distinctness and uniformity testing using SLAF-sequencing approach in soybean. *Genes*. 2020;**11**:175.
12. Zhang D, Vega FE, Infante F et al. Accurate differentiation of green beans of Arabica and Robusta coffee using nanofluidic array of single nucleotide polymorphism (SNP) markers. *J AOAC Int*. 2020;**103**:315–24.
13. Kishor DS, Song WH, Noh Y et al. Development of SNP markers and validation assays in commercial Korean melon cultivars, using genotyping-by-sequencing and Fluidigm analyses. *Sci Hortic*. 2020;**263**:1–9.
14. Nguyen NN, Kim M, Jung JK et al. Genome-wide SNP discovery and core marker sets for assessment of genetic variations in cultivated pumpkin (*Cucurbita* spp.). *Hortic Res*. 2020; **7**:121.
15. Kim M, Jung JK, Shim EJ et al. Genome-wide SNP discovery and core marker sets for DNA barcoding and variety identification in commercial tomato cultivars. *Sci Hortic*. 2021;**276**:1–7.
16. Park G, Choi Y, Jung JK et al. Genetic diversity assessment and cultivar identification of cucumber (*Cucumis sativus* L.) using the Fluidigm single nucleotide polymorphism assay. *Plan Theory*. 2021;**10**:395.
17. Riar DS, Rustgi S, Burke IC et al. EST-SSR development from 5 *Lactuca* species and their use in studying genetic diversity among *L. serriola* biotypes. *J Hered*. 2011;**102**:17–28.
18. Simko I. Development of EST-SSR markers for the study of population structure in lettuce (*Lactuca sativa* L.). *J Hered*. 2009;**100**: 256–62.
19. Uwimana B, D'Andrea L, Felber F et al. A Bayesian analysis of gene flow from crops to their wild relatives: cultivated (*Lactuca sativa* L.) and prickly lettuce (*L. serriola* L.) and the recent expansion of *L. serriola* in Europe. *Mol Ecol*. 2012;**21**:2640–54.
20. Rauscher G, Simko I. Development of genomic SSR markers for fingerprinting lettuce (*Lactuca sativa* L.) cultivars and mapping genes. *BMC Plant Biol*. 2013;**13**:11–1.
21. Hong JH, Kwon YS, Choi KJ et al. Identification of lettuce germplasms and commercial cultivars using SSR markers developed from EST. *Korean J Hort Sci Technol*. 2013;**31**:772–81.
22. Sochor M, Jemelková M, Doležalová I. Phenotyping and SSR markers as a tool for identification of duplicates in lettuce germplasm. *Czech J Genet Plant Breed*. 2019;**55**:110–9.
23. Rui S, Qi G, Shuangxi F et al. Analysis of genetic diversity in purple lettuce (*Lactuca sativa* L.) by SSR markers. *Pak J Bot*. 2020;**52**: 181–96.
24. Choi SP, Sim SC, Hong JH et al. Genetic characterisation of commercial Chinese cabbage varieties using SSR markers. *Seed Sci Technol*. 2016;**44**:595–608.
25. Kong QS, Liu Y, Xie JJ et al. Development of simple sequence repeat markers from de novo assembled transcriptomes of pumpkins. *Plant Mol Biol Rep*. 2020;**38**:130–6.
26. Hong JH, Kwon YS, Mishra RK et al. Construction of EST-SSR databases for effective cultivar identification and their applicability to complement for lettuce (*Lactuca sativa* L.) distinctness test. *Am J Plant Sci*. 2015;**06**:113–25.
27. Raulier P, Maudoux O, Notté C et al. Exploration of genetic diversity within *Cichorium endivia* and *Cichorium intybus* with focus on the gene pool of industrial chicory. *Genet Resour Crop Evol*. 2016;**63**:243–59.
28. Nadeem MA, Nawaz MA, Shahid MQ et al. DNA molecular markers in plant breeding: current status and recent advancements in genomic selection and genome editing. *Biotechnol Biotechnol Equip*. 2018;**32**:261–85.
29. Mammadov J, Aggarwal R, Buyyarapu R et al. SNP markers and their impact on plant breeding. *Int J Plant Genomics*. 2012;**2012**: 1–11.
30. Plomion C, Bartholome J, Lesur I et al. High-density SNP assay development for genetic analysis in maritime pine (*Pinus pinaster*). *Mol Ecol Resour*. 2016;**16**:574–87.
31. Truong HT, Marcos Ramos A, Yalcin F et al. Sequence-based genotyping for marker discovery and codominant scoring in germplasm and populations. *PLoS One*. 2012;**7**:1–9.
32. Kwon S, Simko I, Hellier B *et al.* 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.
33. Park S, Kumar P, Shi A et al. Population genetics and genome-wide association studies provide insights into the influence of selective breeding on genetic variation in lettuce. *Plant Genome*. 2021;**14**:1–12.
34. El-Esawi MA. Molecular genetic markers for assessing the genetic variation and relationships in *Lactuca* germplasm. *Annu Res Rev Biol*. 2015;**8**:1–13.
35. Kandel JS, Peng H, Hayes RJ et al. Genome-wide association mapping reveals loci for shelf life and developmental rate of lettuce. *Theor Appl Genet*. 2020;**133**:1947–66.
36. Egan AN, Schlueter J, Spooner DM. Applications of next-generation sequencing in plant biology. *Am J Bot*. 2012;**99**:175–85.
37. Thomson MJ. High-throughput SNP genotyping to accelerate crop improvement. *Plant Breed Biotechnol*. 2014;**2**:195–212.
38. Kumar, S., Banks, T. W. & Cloutier, S. SNP discovery through next-generation sequencing and its applications. *Int J Plant Genomics* 2012, 831460 (**2012**), 1, 15.
39. D'Agostino N, Tripodi P. NGS-based genotyping, high-throughput phenotyping and genome-wide association studies laid the foundations for next-generation breeding in horticultural crops. *Diversity*. 2017;**9**:38.

40. Le Nguyen K, Grondin A, Courtois B et al. Next-generation sequencing accelerates crop gene discovery. *Trends Plant Sci*. 2019;**24**:263–74.
41. Deschamps S, Llaca V, May GD. Genotyping-by-sequencing in plants. *Biology*. 2012;**1**:460–83.
42. Poland JA, Rife TW. Genotyping-by-sequencing for plant breeding and genetics. *Plant Genome*. 2012;**5**:92–102.
43. Chung YS, Choi SC, Jun TH et al. Genotyping-by-sequencing: a promising tool for plant genetics research and breeding. *Hortic Environ Biotechnol*. 2017;**58**:425–31.
44. Lee KJ, Lee JR, Sebastin R et al. Genetic diversity assessed by genotyping by sequencing (GBS) in watermelon germplasm. *Genes*. 2019;**10**:822.
45. Phan NT, Trinh LT, Rho MY et al. Identification of loci associated with fruit traits using genomewide single nucleotide polymorphisms in a core collection of tomato (*Solanum lycopersicum* L.). *Sci Hortic*. 2019;**243**:567–74.
46. Palumbo F, Qi P, Pinto VB et al. 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.
47. Pereira-Dias L, Vilanova SA, Prohens J et al. Genetic diversity, population structure, and relationships in a collection of pepper (*Capsicum* spp.) landraces from the Spanish Centre of diversity revealed by genotyping-by-sequencing (GBS). *Hortic Res*. 2019;**6**:54.
48. Reuter JA, Spacek DV, Snyder MP. High-throughput sequencing technologies. *Mol Cell*. 2015;**58**:586–97.
49. Slatko BE, Gardner AF, Ausubel FM. Overview of next-generation sequencing technologies. *Curr Protoc Mol Biol*. 2018;**122**:e59.
50. Wang J, Lin M, Crenshaw A et al. High-throughput single nucleotide polymorphism genotyping using nanofluidic dynamic arrays. *BMC Genomics*. 2009;**10**:561.
51. Kim H, Yoon JB, Lee J. Development of Fluidigm SNP type genotyping assays for marker-assisted breeding of chili pepper (*Capsicum annuum* L.). *Hortic Sci Technol*. 2017;**35**:465–79.
52. Choi SR, Heon O, S, Dhandapani, V. et al. Development of SNP markers for marker-assisted breeding in Chinese cabbage using Fluidigm genotyping assays. *Hortic Environ Biotechnol*. 2020;**61**: 327–38.
53. Seo J, Lee G, Jin Z et al. Development and application of *indica–japonica* SNP assays using the Fluidigm platform for rice genetic analysis and molecular breeding. *Mol Breeding*. 2020;**40**:39.
54. Lin Y, Yu W, Zhou L et al. Genetic diversity of oolong tea (*Camellia sinensis*) germplasm based on the nanofluidic array of single-nucleotide polymorphism (SNP) markers. *Tree Genet Genomes*. 2020;**16**:3.
55. Yoo K, Jang S. Intraspecific relationships of *Lactuca sativa* var. *capitata* cultivars based on RAPD analysis. *Korean J Hort Sci Tech*. 2003;**21**:273–8.
56. Sharma S, Kumar P, Gambhir G et al. Assessment of genetic diversity in lettuce (*Lactuca sativa* L.) germplasm using RAPD markers. *3. Biotech*. 2018;**8**:9.
57. Yang TJ, Jang SW, Kim WB. Genetic relationships of *Lactuca* spp. revealed by RAPD, inter SSR, AFLP and PCR-RFLP analyses. *J Crop Sci Biotechnol*. 2007;**10**:27–32.
58. Al-Redhaimam K, Motawei MI, Shinawy M et al. Assessment of genetic variation and presence of nitrate reductase gene (NR) in different lettuce genotypes using PCR-based markers. *J Food Agric Environ*. 2005;**3**:134–6.
59. Lebeda A, Kristkova E, Kitner M et al. Wild *Lactuca* species, their genetic diversity, resistance to diseases and pests, and exploitation in lettuce breeding. *Eur J Plant Pathol*. 2014;**138**:597–640.
60. Zhang L, Su W, Tao R et al. RNA sequencing provides insights into the evolution of lettuce and the regulation of flavonoid biosynthesis. *Nat Commun*. 2017;**8**:2264.
61. He J, Zhao X, Laroche A et al. Genotyping-by-sequencing (GBS), an ultimate marker-assisted selection (MAS) tool to accelerate plant breeding. *Front Plant Sci*. 2014;**5**:484.
62. Kerbiriou PJ, Maliepaard CA, Stomph TJ et al. Genetic control of water and nitrate capture and their use efficiency in lettuce (*Lactuca sativa* L.). *Front Plant Sci*. 2016;**7**:343.
63. Pelgrom AJ, Eikelhof J, Elberse J et al. Recognition of lettuce downy mildew effector BLR38 in *Lactuca serriola* LS102 requires two unlinked loci. *Mol Plant Pathol*. 2019;**20**:240–53.
64. Thiel T, Michalek W, Varshney RK et al. Exploiting EST databases for the development and characterization of gene-derived SSR-markers in barley (*Hordeum vulgare* L.) *Theor*. *Theor Appl Genet*. 2003;**106**:411–22.
65. Meszaros K, Karsai I, Kuti C et al. Efficiency of different marker systems for genotype fingerprinting and for genetic diversity studies in barley (*Hordeum vulgare* L.) *S. S Afr J Bot*. 2007;**73**: 43–8.
66. Singh RK, Sharma RK, Singh AK et al. Suitability of mapped sequence tagged microsatellite site markers for establishing distinctness, uniformity and stability in aromatic rice. *Euphytica*. 2004;**135**:135–43.
67. Yang X, Ren R, Ray R et al. A genetic diversity and population structure of core watermelon (*Citrullus lanatus*) genotypes using DArTseq-based SNPs. *Plant Genet Resour*. 2016;**14**:226–33.
68. Chen W, Hou L, Zhang Z et al. Genetic diversity, population structure, and linkage disequilibrium of a core collection of *Ziziphus jujuba* assessed with genome-wide SNPs developed by genotyping-by-sequencing and SSR markers. *Front Plant Sci*. 2017;**8**:575.
69. Elshire RJ, Glaubitz JC, Sun Q et al. A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. *PLoS One*. 2011;**6**:1–10.
70. Martin M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet J*. 2011;**17**:10–2.
71. Cox MP, Peterson DA, Biggs PJ. SolexaQA: at-a-glance quality assessment of Illumina second-generation sequencing data. *BMC bioinformatics*. 2010;**11**:485.
72. Li H, Durbin R. Fast and accurate short read alignment with burrows-wheeler transform. *Bioinformatics*. 2009;**25**: 1754–60.
73. Kim JE, Oh SK, Lee JH et al. Genome-wide SNP calling using next generation sequencing data in tomato. *Mol Cells*. 2014;**37**:36–42.
74. Li H, Handsaker B, Wysoker A et al. The sequence alignment/map format and SAMtools. *Bioinformatics*. 2009;**25**: 2078–9.
75. Tamura K, Stecher G, Peterson D et al. MEGA6: molecular evolutionary genetics analysis version 6.0. *Mol Biol Evol*. 2013;**30**: 2725–9.
76. Zheng X, Levine D, Shen J et al. A high-performance computing toolset for relatedness and principal component analysis of SNP data. *Bioinformatics*. 2012;**28**:3326–8.
77. Pritchard JK, Stephens M, Donnelly P. Inference of population structure using multilocus genotype data. *Genetics*. 2000;**155**: 945–59.
78. Evanno G, Regnaut S, Goudet J. Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Mol Ecol*. 2005;**14**:2611–20.
79. Rohlf, F. J. NTSYSpc (Numerical Taxonomy and Multivariate Analysis System). Version 2.2, Exeter Software, Applied Biostatistics Inc., New York, USA. (2005).