Development of SNP Markers for the Identification of Commercial Korean Watermelon Cultivars Using Fluidigm Genotyping Analysis

Jun-Young Park[1†], Yoon Jeong Jang[1†], Jin-Kee Jung[2], Eun-Jo Shim[2], Sung-Chr Sim[3], Sang-Min Chung[4], and Gung Pyo Lee[1*] (iD)

[1]Department of Plant Science and Technology, Chung-Ang University, Anseong 17546, Korea
[2]Seed Testing and Research Center, Korea Seed & Variety Service, Gimcheon 39660, Korea
[3]Department of Bioresources Engineering, Sejong University, Seoul 05006, Korea
[4]Department of Life Sciences, Dongguk University, Seoul 04620, Korea

*Corresponding author: gplee@cau.ac.kr

[†]These authors contributed equally to this work.

Abstract

Molecular markers based on simple sequence repeats (SSRs) from expressed sequence tags (ESTs) have been used to identify the registered commercial cultivars in watermelon (*Citrullus lanatus*). To characterize diversity in watermelon cultivars, molecular markers based on genome-wide single nucleotide polymorphisms (SNPs) are needed. Here, we used Fluidigm genotyping for the development of a core set of SNPs to differentiate Korean watermelon commercial cultivars. A candidate subset of SNPs was discovered by genotyping-by-sequencing using 48 F_1 cultivars. The cultivars could be divided into three subpopulations with 97.5% similarity by unweighted pair group method with arithmetic mean (UPGMA) clustering based on 2,300 SNPs. After filtering loci with a high polymorphism information content (PIC), a subset of 238 SNPs was selected and analyzed in 92 F_1 cultivars on the Fluidigm genotyping system. Among these, 141 polymorphic, bi-allelic SNPs were obtained. To evaluate genetic diversity in 92 cultivars, a principal component analysis (PCA) and hierarchical clustering analysis were performed. The first two axes of the PCA explained only 30.5% of the total variance; however, the 92 cultivars clustered into three subgroups with similar sizes. Approximately 90% of the subgroups remained unchanged in the UPGMA clustering analysis. In addition, when the subset of 141 SNPs was filtered with a threshold PIC of 0.36, a core set of 96 SNP markers were obtained, without identical clustering results. The core set of SNP markers can be used for the identification of commercial Korean watermelon cultivars and for the protection of proprietary rights for new cultivar development.

Additional key words: F_1 hybrids, genetic relationship, genotyping-by-sequencing, PCA, population structure

Introduction

Watermelon (*Citrullus lanatus*, $2n = 2x = 22$) is an economically important horticultural crop in the Cucurbitaceae family. Watermelon belongs to a xerophytic genus and originated in northeastern

Africa 5,000 years ago. Watermelon has been cultivated for over 4,000 years. Cultivated watermelons, so-called desert watermelons, have many desirable traits, such as sweet taste or nutritional value. Thus, genetic diversity tends to be low; however, variation among watermelon accessions has not been studied extensively (Guo et al., 2019). Watermelon breeders have identified variation in consumer preferences among countries. For example, in East Asia, there is a preference for small-sized fruits and thin rinds. In contrast, consumers in the United States tend to prefer large, oblong watermelons with a thick rind (Wu et al., 2019).

According to the FAO, global watermelon production was about 104 million tons. In South Korea, 535,000 tons of watermelon were produced in 2018 (FAO, 2018). Watermelon is one of the popular horticulture crops and is cultivated in central and southern regions of South Korea. The priorities for watermelon breeders in Korea are improving fruit shape and texture, sugar content, and biotic and abiotic stress resistance.

In an analysis of genetic polymorphisms in watermelon cultivars and Plant Introduction (PI) accessions using RAPD markers, cultivated watermelon showed high similarity at the genetic level (92 – 99%) (Levi et al., 2001). Another analysis of 49 watermelon varieties using a simple sequence repeat (SSR) marker approach (Kwon et al., 2010) revealed that SSR markers are not suitable for commercial watermelons owing to the limited marker number and low variation. However, they also used EST-SSR and genomic BAC library-based SSR markers to identify highly polymorphic markers in 49 cultivars.

Molecular markers have a number of advantages in plant breeding. They provide an easy way to identify the genetic basis of valuable traits and to predict breeding results quickly and precisely (Nasab et al., 2020). The development of fast and accurate next-generation sequencing (NGS) technology has made a significant contribution to plant breeding. Single nucleotide polymorphisms (SNPs) provide comprehensive genome-wide data, with an even distribution across the whole genome; they can be used for high-throughput genotyping assays as a tool in breeding programs (Ganal et al., 2009; Mammadov et al., 2012). Genotyping-by-sequencing (GBS) is a powerful, cost-effective NGS-based approach that uses restriction enzymes to collect only digested genomic regions, reducing genomic complexity. Accordingly the GBS approach can be used for SNP discovery, high-throughput genotyping, and linkage mapping in various plants (Poland and Rife, 2012; Sonah et al., 2013; Peterson et al., 2014).

In this study, we analyzed the genetic diversity in domestic commercial watermelon cultivars and determined a core set of SNPs to differentiate among the cultivars. In particular, we adopted GBS and Fluidigm genotyping to discover genome-wide SNPs. Robust core-SNP markers were selected based on analyses of molecular diversity and genetic relationships.

Materials and Methods

Plant Materials and DNA Extraction

A total of 100 F1 watermelon cultivars were obtained from the Korea Seed and Variety Service (KSVS), Republic of Korea (Table 1). Among F_1 cultivars, 48 and 92 cultivars were used for GBS and Fluidigm genotyping, respectively. Five seedlings of each accession were grown in a growth room at $24 \pm 1°C$. Young leaves of 3-week-old plants were sampled and ground in liquid nitrogen. Total genomic DNA was extracted using the DNeasy Plant Pro Kit (Qiagen, Germantown,

Table 1. Watermelon F$_1$ commercial cultivars used for GBS and a Fluidigm analysis

Sample	Cultivar	Company	Analysis type
WM01	Speedggul	Nongwoo Bio	GBS
WM02	Lycosweet_2ho	Partner-Seeds	GBS
WM03	Smartggul	Asia Seeds	GBS
WM04	SeedlessPlusggul	Nongwoo Bio	GBS
WM05	SS_ggul_plus	Samsung Seeds	GBS
WM06	Blackbeta	I-Green	GBS
WM07	Sugarwon	Farm Hannong	GBS
WM08	Gangnamggul	Yiseo	GBS
WM09	Dalgonaggul	Syngenta Seeds	GBS, Fluidigm
WM10	Lycofresh_2ho	Partner Seeds	GBS, Fluidigm
WM11	Gamsooggul	Hyundae Seeds	GBS, Fluidigm
WM12	Hwansangggul	Nongwoo Bio	GBS, Fluidigm
WM13	Santaggul	Nongwoo Bio	GBS, Fluidigm
WM14	12weol	Koregon	GBS, Fluidigm
WM15	Speedplusggul	Nongwoo Bio	GBS, Fluidigm
WM16	Wonderfulggul	Nongwoo Bio	GBS, Fluidigm
WM17	Gangta	RDA, Korea	GBS, Fluidigm
WM18	Hwanhiggul	Dongbu Hannong	GBS, Fluidigm
WM19	Jinhansambokggul	Dongbufarm Hannong	GBS, Fluidigm
WM20	Heukdongja	Kwonnong Seeds	GBS, Fluidigm
WM21	Aceggul	Seminis Korea	GBS, Fluidigm
WM22	Jeoktoma	Dongbufarm Hannong	GBS, Fluidigm
WM23	Seolgang	Asia Seeds	GBS, Fluidigm
WM24	Sparkplus	Jangchoon Seeds	GBS, Fluidigm
WM25	HoneyQ_alpha	Asia Seeds	GBS, Fluidigm
WM26	Jijonggul	Dongbu Hitech	GBS, Fluidigm
WM27	Onsesang	Joongang Seeds	GBS, Fluidigm
WM28	Heukmi	Samsung Seeds	GBS, Fluidigm
WM29	Crimsonwave	Dongbufarm Hannong	GBS, Fluidigm
WM30	Dangdanghan	Dongbufarm Hannong	GBS, Fluidigm
WM31	Semiggul	Sakada Korea	GBS, Fluidigm
WM32	Hambakggul	Danong	GBS, Fluidigm
WM33	Hwangryongpo	Samsung Seeds	GBS, Fluidigm
WM34	KD617	Koregon	GBS, Fluidigm
WM35	Mamedeun	Dongbufarm Hannong	GBS, Fluidigm
WM36	Manbokggul	Jeiljongmyonongsan	GBS, Fluidigm
WM37	Miniblackcall	Dongbufarm Hannong	GBS, Fluidigm
WM38	Hero	Dongbufarm Hannong	GBS, Fluidigm
WM39	Sinus	Yiseo	GBS, Fluidigm
WM40	Gangryeoksambokggul	Seminis Korea	GBS, Fluidigm
WM41	Megaspeedggul	Nongwoo Bio	GBS, Fluidigm
WM42	Sinseolgang102	Asia Seeds	GBS, Fluidigm
WM43	Bakmagold	Koregon	GBS, Fluidigm
WM44	Nuneddineggul	Syngenta Seeds	GBS, Fluidigm
WM45	Uriggul	Nongwoo Bio	GBS, Fluidigm
WM46	Blackking	Nongwoo Bio	GBS, Fluidigm
WM47	Creamstar	Dongbufarm Hannong	GBS, Fluidigm
wm96	Seolwhaggul	Koregon	GBS, Fluidigm
WM49	Gigachanggul	Sangrok Seed Farm	Fluidigm
WM50	Sunnyjunior	Kyeongsin Seeds	Fluidigm

Table 1. Watermelon F_1 commercial cultivars used for GBS and a Fluidigm analysis (Continued)

Sample	Cultivar	Company	Analysis type
WM51	Apul	Syngenta Korea	Fluidigm
WM52	Aehong	Kwonnong Seeds	Fluidigm
WM53	Chamjoeun	Taesung Seeds	Fluidigm
WM54	Chosaengheukmi	Samsung Seeds	Fluidigm
WM55	Eliteggul	Sinnong	Fluidigm
WM56	Blackboss	Danong	Fluidigm
WM57	Dandanhan_blackyellow	Daeyeon breeding Inst.	Fluidigm
WM58	Renaplus	Dongbufarm Hannong	Fluidigm
WM59	Sijeogeun	Partner Seeds	Fluidigm
WM60	WM38	Chun Seeds	Fluidigm
WM61	Minigold	Friends & Biz	Fluidigm
WM62	Jalnanggul	Jenong	Fluidigm
WM63	Goodchoice	Farm Hannong	Fluidigm
WM64	Chookbok	Namnongwonye	Fluidigm
WM65	Palio	Hyundae Seeds	Fluidigm
WM66	Orangesugar	Partner Seeds	Fluidigm
WM67	Blackbeta	I-Green	Fluidigm
WM68	Blackinred	Sangrok Seed Farm	Fluidigm
WM69	Blackinyellow	Sangrok Seed Farm	Fluidigm
WM70	Dynamicggul	Sinnong	Fluidigm
WM71	Sweetbongbong	Kyeongwon Seeds	Fluidigm
WM72	Yeorumen	Asia Seeds	Fluidigm
WM73	Luckyplus	Daeilbio Seeds	Fluidigm
WM74	Seongboggul	Sangrok Seed Farm	Fluidigm
WM75	Superballgold	Friends & Biz	Fluidigm
WM76	Managgul	Koregon	Fluidigm
WM77	Dandanhan_blackred	Daeyeon breeding Inst.	Fluidigm
WM78	Babybox	Jinjong Bio	Fluidigm
WM79	Smallhoney	Daerim Seeds	Fluidigm
WM80	Nokboseok	Daeilbio Seeds	Fluidigm
WM81	Habokggul	Jangchoon Seeds	Fluidigm
WM82	Eomjichukggul	Sangrok Seed Farm	Fluidigm
WM83	Onlyyou	Friends & Biz	Fluidigm
WM84	Choonsangrokggul	Sangrok Seed Farm	Fluidigm
WM85	Dandanhan_miniplus	Daeyeon breeding Inst.	Fluidigm
WM86	Minimi	Bayer AG	Fluidigm
WM87	Royalblack	Jinjong Bio	Fluidigm
WM88	Bakangsggul	Sinnong	Fluidigm
WM89	Epyeonhanred	Dana Seeds	Fluidigm
WM90	Hiddencard	Yuan Seeds	Fluidigm
WM91	Jeilblack	Jeiljongmyonongsan	Fluidigm
WM92	Bellocheggul	Sinnong	Fluidigm
WM93	Heukok	NongHyeb	Fluidigm
WM94	Babylemon	Jinjong Bio	Fluidigm
WM95	Sangnongbok	Saengnong	Fluidigm
WM96	Chamhanbok	KMS Seeds	Fluidigm
WM97	Joeunheukggul	Jeiljongmyonongsan	Fluidigm
WM98	Goodchance	Farm Hannong	Fluidigm
WM99	AW1508	Asia Seeds	Fluidigm
WM100	Arongbok	Green Heart Bio	Fluidigm

MD, USA) according to the manufacturer's protocol. The quality of the DNA was checked using a DeNovix DS-11 spectrophotometer (DeNovix, Wilmington, DE, USA) and the amount of double-stranded DNA was measured using the Quant-iT^{TM} Kit, according to the manufacturer's instructions (Themo Fisher Scientific, Waltham, MA, USA).

GBS Library Construction and Illumina Sequencing

To construct the GBS library, each genomic DNA sample was digested with *Ape*KI restriction enzyme (NEB, Ipswich, MA, USA) and ligated to barcoded and common adaptors with T4 DNA ligase. The ligated DNA fragments were pooled and cleaned using the QIAquick PCR purification kit (Qiagen). The cleaned DNA was amplified to prepare the GBS library, and the quantity and quality of which were measured using an Agilent Bioanalyzer (Agilent Technologies, Santa Clara, CA, USA). The GBS library was sequenced with the Illumina HiSeq 2500 system (Illumina, San Diego, CA, USA) by Macrogen (South Korea).

Variant Calling and Marker Development

The raw reads from the GBS library were analyzed using the TASSEL-GBS pipeline (https://bitbucket.org/tasseladmin/tassel-5-source/wiki/Tassel5GBSv2Pipeline). Good barcoded read-tags were mapped on the watermelon reference sequences (97103v2) of the Cucurbit Genomes Database (CuGenDB; http://cucurbitgenomics.org/) using the Burrows-Wheeler Aligner v0.6.1-r104 (BWA) (Li and Durbin, 2010). The resulting SNPs were filtered under the conditions of minor allele frequency (MAF) > 5%, missing data < 20%, and minimum depth > 10 to produce a variant calling format file. The functional effects of the SNPs were predicted using the snpEff toolbox (Cingolani et al., 2012). The filtered SNPs from 48 F_1 cultivars were selected based on the polymorphism information content (PIC) (Bostein et al., 1980) value to obtain a core marker set, which was applied to the Fluidigm assay for discriminating 92 watermelon cultivars. The PIC value was calculated using PowerMarker V3.25 (Liu and Muse, 2005). The Fluidigm assay was performed using the Fluidigm Juno system (Fluidigm, San Francisco, CA, USA). Allele-specific primers were used for PCR amplification and SNP genotyping using the Juno 96.96 Dynamic Array IFC (Integrated Fluidic Circuit), and genotypes were analyzed using Fluidigm SNP Genotyping Analysis v4.5.1.

Population Structure and Genetic Relationships

The population structure was evaluated based on a Bayesian model-based clustering method using STRUCTURE v2.3.4 (Pritchard et al., 2000). The K values were set from 1 to 10 with a burn-in period of 10,000 iterations. Markov chain Monte Carlo (MCMC) runs were carried out with 100,000 iterations per run. The number of replications for each K was set to 10. The best K value was predicted according to the Evanno method using STRUCTURE HARVESTER v0.6.94 (Earl et al., 2012) by calculating the ΔK. A principal component analysis (PCA) was performed using the prcomp function and results were visualized using the ggplot2 package (Wickham, 2009) in R. A hierarchical clustering analysis was performed using the poppr package (Kamvar et al., 2014) implemented in R. The poppr package was used to estimate Nei's genetic distances (Nei, 1978) between F_1 cultivars, and a hierarchical clustering analysis was conducted by the unweighted pair group method with arithmetic means (UPGMA). The UPGMA dendrogram was generated using the dendextent package (Galili, 2015).

Results

GBS Analysis

A GBS analysis of 48 watermelon F_1 commercial cultivars, registered in the KSVS (Cimcheon, Republic of Korea), produced 204.1 million reads, which were analyzed in the Tassel-GBS pipeline (Table 2). A total of 18,888 bi-allelic SNPs were mapped to SNP positions with the good barcoded reads. After filtering the SNPs with thresholds of MAF >5% and missing data <10%, we obtained 9,397 SNPs. These SNPs were distributed across the reference genome (Suppl. Fig. 1). Then, we obtained 2,300 SNPs with >10× depth coverage, including one SNP variant for every 178,595 bases of the watermelon reference genome (Table 3).

Population Structure and Genetic Relationships

To identify subgroups of the 48 F_1 commercial cultivars, a population structure analysis was performed using a Bayesian clustering approach as implemented in STRUCTURE (Fig. 1). The optimal number of subpopulations was 3, as determined by the calculation of ΔK (Fig. 1A). When considering the unknown parental populations, each subpopulation was ambiguously assigned to specific subspecies or varieties. To further understand the genetic relatedness and diversity of the cultivars, we built a tree based on the UPGMA algorithm with 2,300 SNPs (Suppl. Fig. 2). Most samples were assigned to group I (similarity −97.5%), and only three cultivars were exclusively clustered to group II ('WM41' and

Table 2. Summary of genotyping-by-sequencing of the 48 F_1 watermelon cultivars

Number of cultivars	48
Total number of reads	204,148,192
Total number of good barcoded reads	102,101,726
Size of all tags	257,680
Size of all SNP positions	18,888
Size of filtered SNPs	9,397

Table 3. Number of variants and variant rate for each chromosome

Chromosome	Length (bp)	Variants (SNPs)	Variant rate (bp)
1	36,935,898	289	127,805
2	37,915,939	150	252,772
3	31,872,261	166	192,001
4	27,110,815	126	215,165
5	35,887,987	174	206,252
6	29,507,460	245	120,438
7	31,939,013	119	268,395
8	28,201,227	139	202,886
9	37,727,573	185	203,932
10	35,099,344	141	248,931
11	30,886,124	299	103,298
Total	363,083,641	2,033	178,595[z]

[z]Mean value of genomic variant rates: one variant for every 178,595 bases in the watermelon genome (97103v2).

'WM37') and group III ('WM42'). The three cultivars were also detected in the population structure analysis without admixture setting K= 2 and little admixture setting K = 3 (Fig. 1B).

Fluidigm SNP Assay and Marker Development

To develop highly polymorphic SNP markers to discriminate cultivars, a subset of 238 SNPs was selected based on PIC > 0.3 and physical distance on each chromosome. These 238 SNPs were tested for the genotyping of newly prepared 92 F_1 cultivars (Table 1) using the Fluidigm Juno system. The success rates of amplification per F_1 cultivar ranged from 83.5% to 100%, with an average of 92.8%. The genotype calling results for four assays are summarized in Fig. 2. A subset of 141 bi-allelic SNPs was obtained, and these SNPs were used for a genetic analysis.

Genetic Analysis and Core Marker Set Development for 92 F_1 Cultivars

For an overview of the 141 SNPs, PCA was applied (Fig. 3). The first two principal components (PC1 and PC2) accounted for 30.5% of the total variance (18.6% and 11.9%, respectively). A cluster analysis using the silhouette test
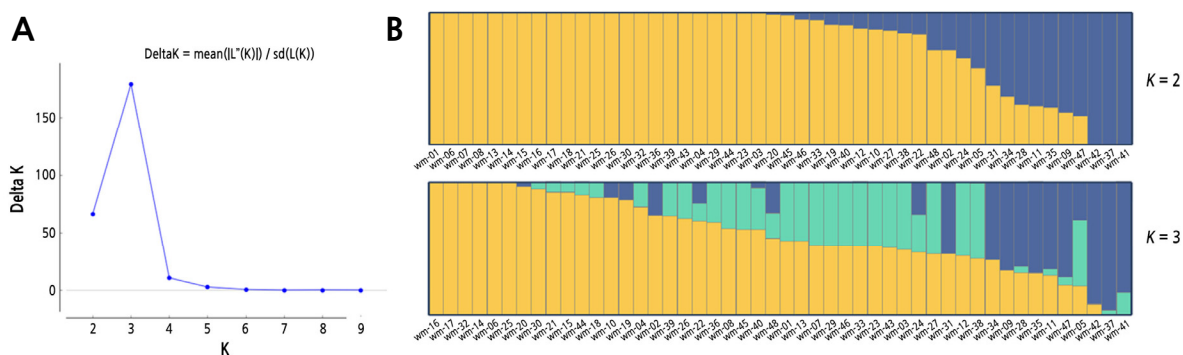
Fig. 1. Population structure analysis of 48 watermelon F_1 cultivars using STRUCTURE based on 2,300 SNPs. (A) Plot of ΔK values from K=1 to 10. (B) Admixture-based substructure model. Each color indicates a presumed ancestral population.

Fig. 2. Scatterplots of the Fluidigm genotyping analysis. A total of 92 samples were analyzed based on biallelic SNPs: wm96_8S21 (left panel) and wm96_9S06 (right panel). Dots are clustered as homozygote allele 1 (red), homozygote allele 2 (green), and heterozygote (blue).

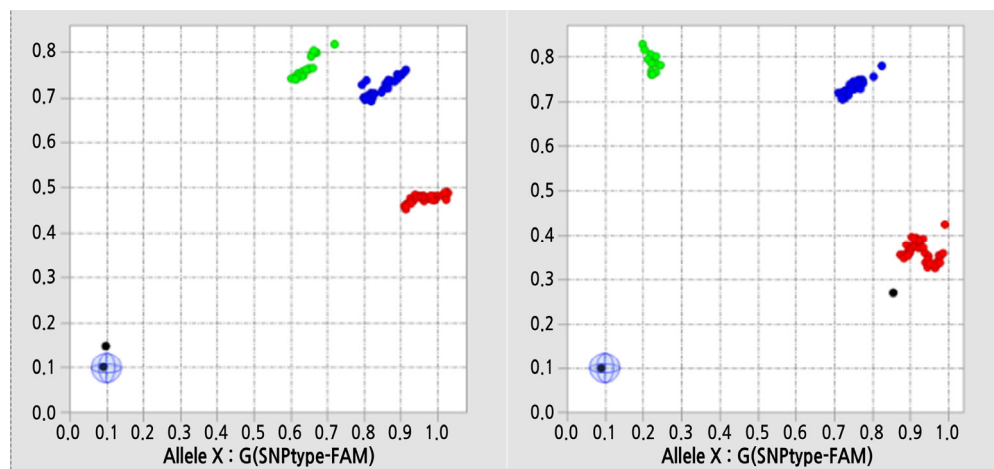resulted in K = 3 clusters, as visualized by scatter plots (Fig. 3). Most of the cultivars were separated and clustered into three subgroups.

Next, with the subset of 141 SNPs, hierarchical clustering was performed using the UPGMA algorithm based on Nei's genetic distances (Fig. 4). The cultivars were separated into three clusters using the cutree function in R: subgroup 1 (n = 30), subgroup 2 (n = 37), and subgroup 3 (n = 23). The cluster number (K = 3) was the same; however, the subgroup membership differed slightly between the PCA and hierarchical clustering methods.
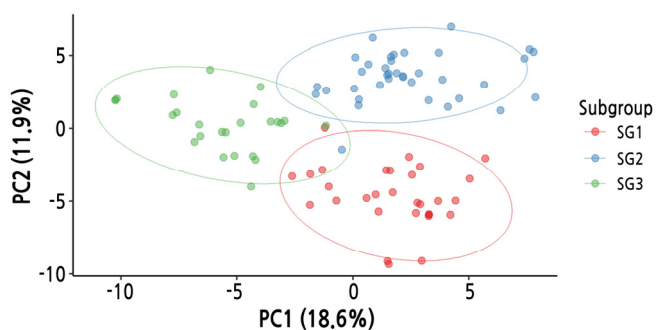
Fig. 3. Two-dimensional principal component analysis (PCA) of 92 watermelon F_1 cultivars. PC1 and PC2 are shown in the scatter plot and subgroups are indicated using colored dots. Each ellipse represents the 95% confidence interval of a subgroup.
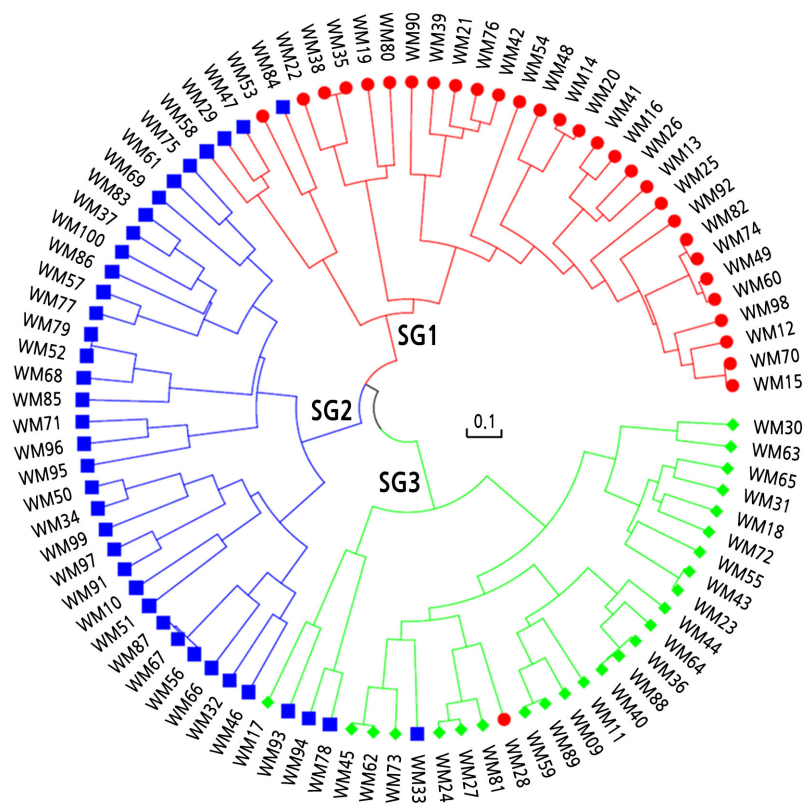
Fig. 4. Dendrogram using the UPGMA algorithm based on Nei's distances. The genetic distance was calculated using 114 SNPs and 92 watermelon cultivars. Subgroups 1, 2, and 3 are shown in red, blue, and green, respectively. Differently colored dots correspond to the subgroups in Fig. 3.

The subset of 141 SNPs was reduced to 96 SNPs for the analysis using the Fluidigm Juno 96.96 system. To prepare these 96 SNPs, a threshold PIC of > 0.36 and the physical position of SNPs were considered. The core set of 96 SNP markers has a heterozygosity per F_1 cultivar of 9.2% to 82.3%, with an average of 48.9%. The average values of MAF, gene diversity, heterozygosity, and PIC were 0.596, 0.473. 0.509, and 0.361, respectively (Suppl. Table 2). The UPGMA dendrogram based on the 96 SNPs did not show any differences in topology from that based on 141 SNPs.

Discussion

In this study, the Fluidigm analysis platform was used as a cost-effective approach for the development of core SNP markers to differentiate commercial watermelon cultivars. First, we obtained variants including SNPs based on genome-wide complexity reduction by selectively sequencing the barcoded fragments digested with *Ape*KI restriction enzymes (Poland and Rife, 2012; He et al., 2014; Jung et al., 2020). With a filtering threshold of >10× depth coverage, we obtained a number of reasonably valid SNPs (2,300 SNPs) for downstream analyses. The SNPs were analyzed based on the annotated transcripts in the reference genome using snpEff (Cingolani et al., 2012). For example, 1,203 of 4,108 transcription-related factors (24.1%) were detected in exonic regions (Suppl. Table 1), consistent with soybean GBS results, in which 20.73% of SNPs resided in exonic regions when using a methylation-sensitive *Ape*KI restriction enzyme (Sonah et al., 2013).

Based on the SNPs, we analyzed the population structure and genetic relationships of 48 commercial watermelon cultivars to understand their molecular diversity. The cultivars were assigned to three subpopulations in the population structure analysis. However, a dendrogram based on genetic distances (Suppl. Fig. 2) revealed little variation in the genetic background, which is a well-known characteristic of watermelon cultivars, which have mainly been bred for a high sugar content and large fruits (Wu et al., 2019; Kim et al., 2021).

Next, we focused on the selection of polymorphic SNPs to establish a core set of markers able to discriminate among domestic watermelon cultivars registered in the KSVS. We determined a subset of 238 SNPs with an even distribution based on physical distance and PIC > 0.3 in a molecular diversity analysis. The SNPs were tested in 92 F_1 cultivars using the Fluidigm Juno system. After further filtering, 141 bi-allelic SNPs were identified. The average heterozygosity was 50.9% between the F_1 cultivars. The genotyping data showed moderate genetic distances, and the SNP markers could be used to test the purity of the registered F_1 cultivars (Kishor et al., 2020). Further, the genotyping data were analyzed by a PCA and hierarchical clustering analysis, in which the 92 cultivars were classified into three subgroups. For most cultivars, consistent results were obtained by the two approaches; however, nine cultivars showed differences in group assignments between the PCA and hierarchical clustering analysis. This inconsistency can be explained by the small number of tested SNPs and narrow genetic background of the tested cultivars. The subset of 141 SNPs was reduced to a core subset of 96 SNPs for an analysis using the Fluidigm Juno 96.96 platform. The dendrogram based on 141 SNPs was identical to the dendrogram based on 96 SNPs (data not shown). With this core SNP set, we could accurately identify the registered 92 F_1 cultivars.

In this study, we reported a core set of 96 SNP markers for the identification of domestically registered 92 F_1 watermelon cultivars. Genetic analyses revealed that the domestic cultivars have low levels of genetic diversity. However, the newly developed SNP markers are useful to differentiate domestic watermelon cultivars by a high-throughput analysis platform, such as the Fluidigm genotype system. In addition, the newly established markers could protect proprietary rights for new cultivars developed by breeders at the molecular level.

Literature Cited

Bostein D, White RL, Skolnick M, Davis RW (1980) Construction of a genetic linkage map in man using restriction fragment length polymorphisms. Am J Hum Genet 32:314-331

Cingolani P, Platts A, Wang le L, Coon M, Nguyen T, Wang L, Land SJ, Lu X, Ruden DM (2012) A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of Drosophila melanogaster strain w1118; iso-2; iso-3. Fly (Austin) 6:80-92. doi:10.4161/fly.19695

Food and Agriculture Organization of the United Nations (FAO) (2018) Statistical databases. Food and Agriculture Organization of the United Nations, http://www.fao.org/faostat/en/#data/QC. Accessed 21 October 2021

Galili T (2015) dendextend: an R package for visualizing, adjusting and comparing trees of hierarchical clustering. Bioinformatics 31:3718-3720. doi:10.1093/bioinformatics/btv428

Ganal MW, Altmann T, Roder MS (2009) SNP identification in crop plants. Curr Opin Plant Biol 12:211-217. doi:10.1016/j.pbi.2008.12.009

Guo S, Zhao S, Sun H, Wang X, Wu S, Lin T, Ren Y, Gao L, Deng Y, et al (2019) Resequencing of 414 cultivated and wild watermelon accessions identifies selection for fruit quality traits. Nat Genet 51:1616-1623. doi:10.1038/s41588-019-0518-4

He J, Zhao X, Laroche A, Lu ZX, Liu H, Li Z (2014) Genotyping-by-sequencing (GBS), an ultimate marker-assisted selection (MAS) tool to accelerate plant breeding. Front Plant Sci 5:484. doi:10.3389/fpls.2014.00484

Jung J, Park G, Oh J, Jung JK, Shim EJ, Chung SM, Lee GP, Park Y (2020) Assessment of the current infraspecific classification scheme in melon (*Cucumis melo* L.) based on genome-wide single nucleotide polymorphisms. Hortic Environ Biotechnol 61:537-547. doi:10.1007/s13580-020-00230-0

Kamvar ZN, Tabima JF, Grunwald NJ (2014) Poppr: an R package for genetic analysis of populations with clonal, partially clonal, and/or sexual reproduction. PeerJ 2:e281. doi:10.7717/peerj.281

Kim M, Jung JK, Shim EJ, Chung SM, Park Y, Lee GP, Sim SC (2021) Genome-wide SNP discovery and core marker sets for DNA barcoding and variety identification in commercial tomato cultivars. Sci Hortic 276:109734. doi:10.1016/j.scienta.2020.109734

Kishor D, Noh Y, Song WH, Lee GP, Jung JK, Shim EJ, Chung SM (2020) Identification and Purity Test of Melon Cultivars and F1 Hybrids Using Fluidigm-based SNP Markers. Hortic Sci Technol 38:686-694

Kwon YS, Oh YH, Yi SI, Kim HY, An JM, Yang SG, Ok SH, Shin JS (2010) Informative SSR markers for commercial variety discrimination in watermelon (*Citrullus lanatus*). Genes Genomics 32:115-122. doi:10.1007/s13258-008-0674-x

Levi A, Thomas CE, Wehner TC, Zhang XP (2001) Low genetic diversity indicates the need to broaden the genetic base of cultivated watermelon. J Am Soc Hortic Sci 36:1096-1101. doi:10.21273/HORTSCI.36.6.1096

Li H, Durbin R (2010) Fast and accurate long-read alignment with Burrows-Wheeler transform. Bioinformatics 26:589-595. doi:10.1093/bioinformatics/btp698

Liu K, Muse SV (2005) PowerMarker: an integrated analysis environment for genetic marker analysis. Bioinformatics 21:2128-2129. doi:10.1093/bioinformatics/bti282

Mammadov J, Aggarwal R, Buyyarapu R, Kumpatla S (2012) SNP markers and their impact on plant breeding. Int J Plant Genomics 2012. doi:10.1155/2012/728398

Nasab MA, Rahimi M, Karatas A, Ercisli S (2020) Sequential path analysis and relationships between fruit yield in watermelon. Pakistan J Agri Sci 57:1425-1430. doi:10.21162/Pakjas/20.200

Nei M (1978) Estimation of average heterozygosity and genetic distance from a small number of individuals. Genetics 89:583-590. doi:10.1093/genetics/89.3.583

Peterson G, Dong Y, Horbach C, Fu YB (2014) Genotyping-By-Sequencing for Plant Genetic Diversity Analysis: A Lab Guide for SNP Genotyping. Diversity 6:665-680. doi:10.3390/d6040665

Poland JA, Rife TW (2012) Genotyping-by-Sequencing for Plant Breeding and Genetics. Plant Genome 5:92-102. doi:10.3835/plantgenome2012.05.0005

Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. Genetics 155. doi:10.1093/genetics/155.2.945

Sonah H, Bastien M, Iquira E, Tardivel A, Légaré G, Boyle B, Normandeau É, Laroche J, Larose S, et al (2013) An improved genotyping by sequencing (GBS) approach offering increased versatility and efficiency of SNP discovery and genotyping. PLoS ONE 8:e54603. doi:10.1371/journal.pone.0054603

Wickham H (2009) ggplot2: elegant graphics for data analysis. Springer, NY, USA. doi:10.1007/978-0-387-98141-3

Wu S, Wang X, Reddy U, Sun HH, Bao K, Gao L, Mao LY, Patel T, Ortiz C, et al (2019) Genome of 'Charleston Gray', the principal American watermelon cultivar, and genetic characterization of 1,365 accessions in the US National Plant Germplasm System watermelon collection. Plant Biotechnol J 17:2246-2258. doi:10.1111/pbi.13136

Supplementary Table 1. Number of SNPs by the variant type and genomic region

| Type		
Type (alphabetical order)	Count	Percentage (%)
Downstream_gene_variant	1,029	24.921
Intergenic_region	731	17.704
Intron_variant	275	6.666
Missense_variant	537	13.006
Splice_acceptor_variant	1	0.024
Splice_region_variant	20	0.484
Start_lost	1	0.024
Stop_gained	28	0.678
Stop_lost	3	0.073
Synonymous_variant	458	11.092
Upstream_gene_variant	1,046	25.333
Region		
Type (alphabetical order)	Count	Percentage (%)
Downstream	1,029	25.049
Exon	1,023	24.093
Intergenic	731	17.795
Intron	265	6.451
Splice_site_acceptor	1	0.024
Splice_site_region	13	0.316
Upstream	1,046	25.463

Supplementary Table 2. Genotype summary of 96-plex watermelon SNP primers based on 92 watermelon F_1 commercial cultivars

| No	Marker	Chr	MAF	GD	Heterozygosity	PIC
1	wm96_1S01	1	0.690	0.428	0.402	0.336
2	wm96_1S04	1	0.549	0.495	0.505	0.373
3	wm96_1S05	1	0.593	0.483	0.396	0.366
4	wm96_1S07	1	0.648	0.456	0.374	0.352
5	wm96_1S11	1	0.527	0.498	0.945	0.374
6	wm96_1S13	1	0.630	0.466	0.478	0.357
7	wm96_1S16	1	0.588	0.485	0.341	0.367
8	wm96_1S17	1	0.516	0.499	0.505	0.375
9	wm96_1S20	1	0.544	0.496	0.363	0.373
10	wm96_1S21	1	0.565	0.491	0.500	0.371
11	wm96_2S02	2	0.505	0.500	0.549	0.375
12	wm96_2S04	2	0.663	0.447	0.457	0.347
13	wm96_2S06	2	0.588	0.485	0.626	0.367
14	wm96_2S13	2	0.637	0.462	0.462	0.355
15	wm96_2S15	2	0.621	0.471	0.429	0.360
16	wm96_2S20	2	0.577	0.488	0.473	0.369
17	wm96_2S22	2	0.566	0.491	0.516	0.371
18	wm96_2S24	2	0.599	0.480	0.538	0.365
19	wm96_2S25	2	0.670	0.442	0.462	0.344
20	wm96_3S02	3	0.614	0.474	0.467	0.362
21	wm96_3S04	3	0.511	0.500	0.473	0.375
22	wm96_3S10	3	0.610	0.476	0.780	0.363
23	wm96_3S12	3	0.632	0.465	0.495	0.357
24	wm96_3S15	3	0.641	0.460	0.500	0.354
25	wm96_3S19	3	0.685	0.432	0.413	0.339
26	wm96_3S22	3	0.511	0.500	0.630	0.375
27	wm96_4S02	4	0.516	0.499	0.946	0.375
28	wm96_4S03	4	0.571	0.490	0.620	0.370
29	wm96_4S06	4	0.522	0.499	0.522	0.375
30	wm96_4S09	4	0.554	0.494	0.891	0.372
31	wm96_4S10	4	0.707	0.415	0.348	0.329
32	wm96_4S12	4	0.713	0.410	0.483	0.326
33	wm96_5S01	5	0.668	0.443	0.337	0.345
34	wm96_5S08	5	0.565	0.491	0.413	0.371
35	wm96_5S11	5	0.554	0.494	0.543	0.372
36	wm96_5S14	5	0.647	0.457	0.511	0.353
37	wm96_5S17	5	0.571	0.490	0.511	0.370
38	wm96_5S18	5	0.571	0.490	0.511	0.370
39	wm96_5S20	5	0.709	0.413	0.582	0.328
40	wm96_5S21	5	0.543	0.496	0.500	0.373
41	wm96_5S23	5	0.783	0.340	0.435	0.282
42	wm96_5S27	5	0.598	0.481	0.565	0.365
43	wm96_6S02	6	0.625	0.469	0.402	0.359
44	wm96_6S03	6	0.690	0.428	0.620	0.336
45	wm96_6S06	6	0.522	0.499	0.500	0.375
46	wm96_6S07	6	0.576	0.488	0.457	0.369
47	wm96_6S08	6	0.598	0.481	0.413	0.365
48	wm96_6S09	6	0.527	0.499	0.467	0.374
49	wm96_6S15	6	0.614	0.474	0.467	0.362
50	wm96_6S20	6	0.543	0.496	0.565	0.373

Supplementary Table 2. Genotype summary of 96-plex watermelon SNP primers based on 92 watermelon F_1 commercial cultivars (Continued)

No	Marker	Chr	MAF	GD	Heterozygosity	PIC
51	wm96_6S23	6	0.663	0.447	0.500	0.347
52	wm96_7S01	7	0.582	0.487	0.424	0.368
53	wm96_7S04	7	0.576	0.488	0.413	0.369
54	wm96_7S05	7	0.522	0.499	0.370	0.375
55	wm96_7S06	7	0.582	0.487	0.467	0.368
56	wm96_7S07	7	0.565	0.491	0.457	0.371
57	wm96_7S08	7	0.527	0.499	0.489	0.374
58	wm96_7S10	7	0.625	0.469	0.489	0.359
59	wm96_7S17	7	0.527	0.499	0.533	0.374
60	wm96_8S06	8	0.663	0.447	0.522	0.347
61	wm96_8S10	8	0.565	0.491	0.870	0.371
62	wm96_8S12	8	0.717	0.405	0.435	0.323
63	wm96_8S13	8	0.717	0.405	0.457	0.323
64	wm96_8S14	8	0.668	0.443	0.511	0.345
65	wm96_8S15	8	0.723	0.401	0.446	0.320
66	wm96_8S16	8	0.592	0.483	0.533	0.366
67	wm96_8S19	8	0.533	0.498	0.435	0.374
68	wm96_8S20	8	0.533	0.498	0.391	0.374
69	wm96_8S22	8	0.511	0.500	0.435	0.375
70	wm96_9S04	9	0.527	0.499	0.554	0.374
71	wm96_9S05	9	0.620	0.471	0.761	0.360
72	wm96_9S06	9	0.652	0.454	0.413	0.351
73	wm96_9S10	9	0.554	0.494	0.478	0.372
74	wm96_9S16	9	0.609	0.476	0.413	0.363
75	wm96_9S18	9	0.538	0.497	0.924	0.374
76	wm96_9S20	9	0.560	0.493	0.467	0.371
77	wm96_9S21	9	0.625	0.469	0.402	0.359
78	wm96_9S25	9	0.511	0.500	0.500	0.375
79	wm96_9S27	9	0.505	0.500	0.467	0.375
80	wm96_10S02	10	0.609	0.476	0.522	0.363
81	wm96_10S05	10	0.720	0.403	0.429	0.322
82	wm96_10S12	10	0.643	0.459	0.516	0.354
83	wm96_10S14	10	0.588	0.485	0.451	0.367
84	wm96_10S15	10	0.681	0.434	0.418	0.340
85	wm96_10S16	10	0.505	0.500	0.396	0.375
86	wm96_11S01	11	0.555	0.494	0.538	0.372
87	wm96_11S02	11	0.632	0.465	0.582	0.357
88	wm96_11S04	11	0.648	0.456	0.505	0.352
89	wm96_11S05	11	0.698	0.422	0.451	0.333
90	wm96_11S08	11	0.677	0.437	0.646	0.342
91	wm96_11S11	11	0.527	0.499	0.533	0.374
92	wm96_11S12	11	0.522	0.499	0.407	0.375
93	wm96_11S13	11	0.659	0.449	0.659	0.348
94	wm96_11S16	11	0.588	0.485	0.473	0.367
95	wm96_11S17	11	0.599	0.480	0.473	0.365
96	wm48_11S18	11	0.637	0.462	0.527	0.355
	Mean		0.596	0.473	0.509	0.361

NA: Allele number; GD: Gene diversity; PIC: Polymorphism information content.
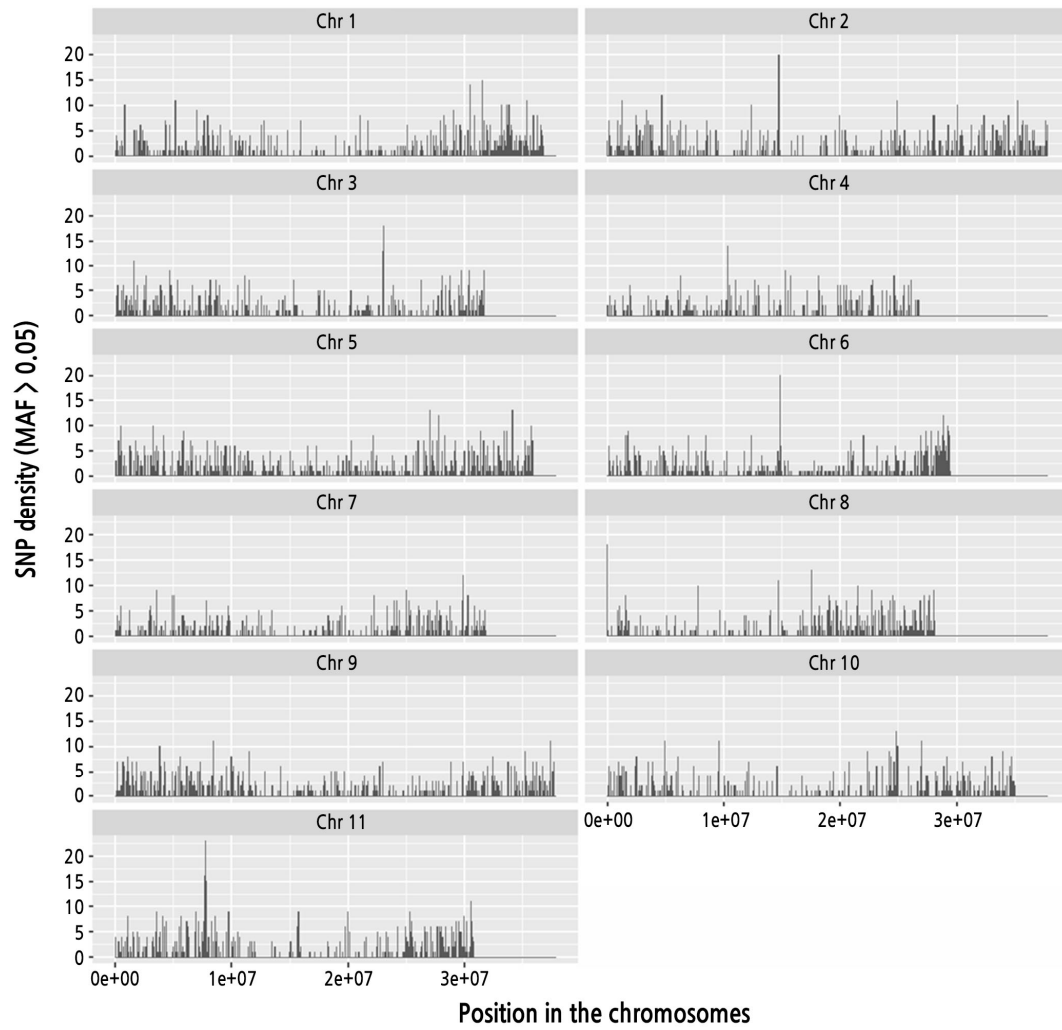
Supplementary Fig. 1. SNP density of 48 watermelon cultivars.
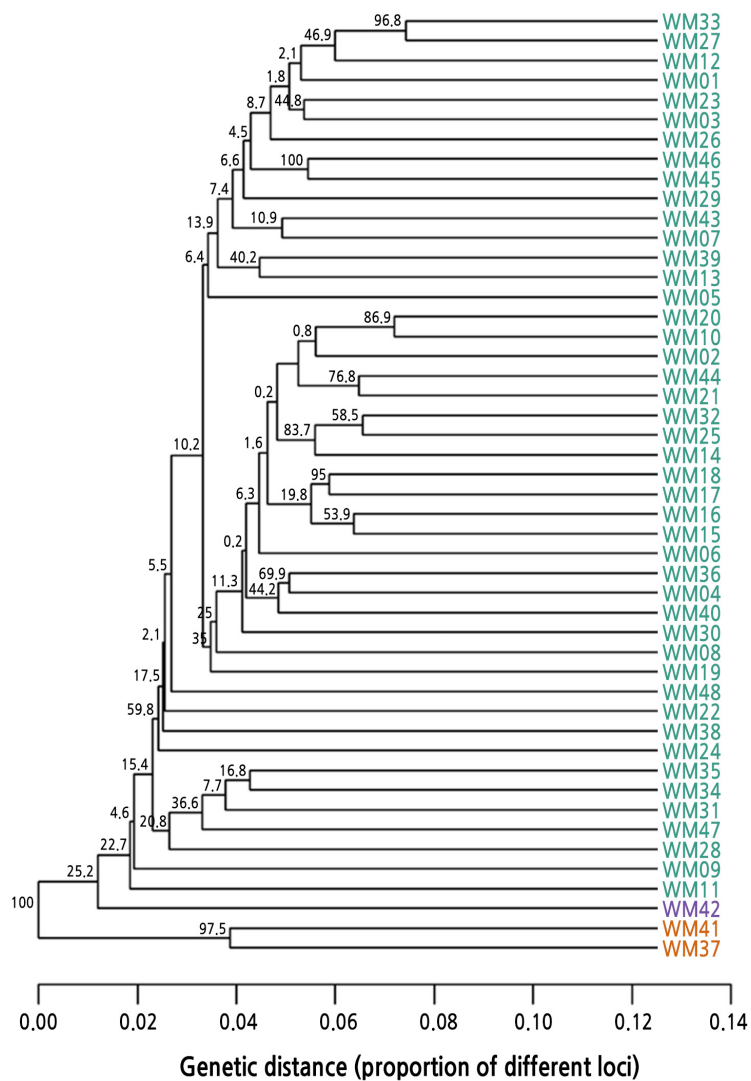
Supplementary Fig. 2. Dendrogram of hierarchical clustering results for the genetic distance analysis.