

Development of Single-Nucleotide Polymorphism Markers for Seed Coat Color in Watermelon

Yoon Jeong Jang^{1†}, Ki Ryong Kim^{1†}, Jisu Ryu¹, Yongjae Kim², Sang-Min Chung³, and Gung Pyo Lee^{1*} 

¹Department of Plant Science and Technology, Chung-Ang University, Anseong 17546, Korea

²Partner Seeds Co., Ltd., Gimje 54324, Korea

³Department of Life Sciences, Dongguk University, Seoul 04620, Korea

*Corresponding author: gplee@cau.ac.kr

[†]These authors contributed equally to this work.

Abstract

Seed coat color affects water absorption, gas diffusion, seed dormancy, seed quality, germination, and seedling appearance. These attributes of watermelon seeds thus necessitate the establishment of watermelon breeding programs that produce desired seed coat colors. However, the hereditary and molecular mechanisms underlying this feature remain unknown. Here, we aimed to identify the genomic region responsible for seed coat color in watermelon and to design single-nucleotide polymorphism (SNP) markers that are closely associated with this characteristic. We crossed two inbred lines, RG-21 (white seed coat, WSC) and Wr-609 (red seed coat, RSC), to generate F₁ hybrids (tan seed coat) and F₂ populations for locus mapping and a SNP marker analysis. Our findings indicated that the F₂ seed coat color is controlled by negative epistasis-expressing WSC. A bulked segregant analysis combined with Illumina whole-genome resequencing revealed a putative causative genomic area for WSC at 5.1–9.7 Mb on chromosome (Chr.) 6 and RSC at 3.6–7.7 Mb on Chr. 5. Based on the four-gene hypothesis, the genotypes of the WSC, RSC, and F₁ hybrids were *Rtwd*, *rtWd*, and *RtWd*, respectively. We then applied 19 high-resolution melting markers to 106 F₂ progeny to identify the causative genomic regions. We created a strongly co-segregated SNP marker for WSC at 6.9–7.1 Mb on Chr. 6 and RSC at 4.2–4.8 Mb on Chr. 5. Our findings serve as a basis for utilizing seed coat color in watermelon breeding programs and for expediting F₁ cultivar generation.

Additional key words: bulked segregant analysis, delta single-nucleotide polymorphism index, high-resolution melting, linkage marker, next-generation sequencing

Introduction

Watermelon (*Citrullus lanatus*; 2n = 2x = 22) is a globally important vegetable crop. In 2022, South Korea reported a production volume of 476,897.99 tons and a cultivated area of approximately 11,346 ha (FAOSTAT, 2022; <https://www.fao.org/faostat/>). In 2023, watermelon production in South Korea amounted to 1,569.9 billion South Korean won (KRW), accounting for 4% of the agricultural

Received: 14 June 2024

Revised: 03 July 2024

Accepted: 05 July 2024

Published online: 25 July 2024



OPEN ACCESS



HORTICULTURAL SCIENCE and TECHNOLOGY
42(4):480-491, 2024
URL: <http://www.hst-j.org>

pISSN : 1226-8763
eISSN : 2465-8588

This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

Copyright©2024 Korean Society for Horticultural Science.

We would like to express our gratitude to Gen-Bank of the National Agrobiodiversity Center (NAC) of the Rural Development Administration (Gimje, Korea) for providing the seeds. This study was supported by the "Cooperative Research Program for Agriculture Science & Technology Development (Project No. PJ014 21302) of the Rural Development Administration of the Republic of Korea." This research was supported by a Chung-Ang University Young Scientist Scholarship in 2020.

Conflicts of Interest

The authors declare no conflicts of interest.

production total. Additionally, seed production reached 593 kg in total (Korean Seed Association; <http://kosaseed.or.kr/>).

Watermelon flesh is high in water and rich in phytochemicals such as lycopene, β -carotene, and ascorbic acid. The lycopene content in watermelons is more than 60% higher than that in tomatoes (Guner and Wehner, 2004). Watermelon contains citrulline, a non-essential amino acid that serves as a source of antioxidants and vasodilators. Citrulline is a by-product generated during the oxidation of arginine and plays a role in protecting blood vessels by delivering nutrients to cells in the cardiovascular system (Rimando and Perkins-Veazie, 2005). Watermelon seeds are also rich in proteins that serve as precursors of antioxidant peptides (Wen et al., 2020). According to previous studies, watermelon seed protein hydrolysates with molecular weights less than 1 kDa exhibit strong antioxidant capabilities (Wen et al., 2019a).

Seed color affects various physiological processes, such as water uptake (Powell, 1989; Mavi, 2010), gas diffusion (West and Harris, 1963), seed dormancy (Baskin et al., 2000), and overall seed quality, owing to certain pigments in the seed coat (Powell, 1989; Mavi, 2010). Seed coat color is linked to the biochemical properties of seeds, including the antioxidant levels and activities (El-Bramawy et al., 2008), and affects the appearance of the flesh as well (Wehner, 2008). Black seeds are attractive when paired with red or canary yellow flesh, whereas near-seedless cultivars show white or light-colored seeds (Wehner, 2008). Among these traits, the coat color of watermelon seeds is of significant economic value, particularly in Asia and West Africa. In these areas, watermelon seeds are cultivated not only for their horticultural qualities but also for their culinary uses in snacks and soups, making seed coat color a critical factor in relation to consumer preferences and marketability (Gusmini et al., 2004; Mavi, 2010).

Watermelon seed coat colors exhibit a wide range of variations, including solid black, dotted black, brown, tan, green, red, white, and black stripes with white (Poole, 1941; Poole, 1944; Paudel et al., 2019). Poole et al. (1941) investigated a genetic model for controlling seed coat color by examining 40 different segregated populations. Their study suggested a three-gene model involving genes *R*, *T*, and *W*, along with a modifier gene, *D*.

With the advent of next-generation sequencing (NGS) technology, sequencing has become a high-throughput, less error-prone and more cost-effective method. Consequently, NGS has been used extensively to identify molecular markers across genomes (Ruangrak et al., 2018). The publicly available watermelon reference genome includes the watermelon cultivar 97103 and the American heirloom variety Charleston Gray. These genomes are available in the Cucurbit Genomics Database, which provides essential resources for molecular marker development, gene mapping, quantitative trait locus (QTL) mapping, and molecular breeding research (Guo et al., 2013; Guo et al., 2019; Wu et al., 2019). A relatively recent advancement in exploring major QTLs is the bulked segregant analysis combining NGS data (BSA-seq; also known as QTL-seq) to identify the QTLs and genetic markers needed for marker-assisted selection (MAS) (Michelmore et al., 1991; Takagi et al., 2013). A major benefit of BSA-seq is that it does not require the genotyping of every individual within a population, thus significantly simplifying the process. In watermelons, BSA-seq was initially used to map a dwarfism locus on chromosome (Chr.) 7 (Dong et al., 2018). This technique has been used to rapidly develop molecular markers of various watermelon traits to enhance breeding efficiency. Additionally, BSA-seq has been used to identify genetic loci in watermelons and other crops, highlighting its versatility and effectiveness (Takagi et al., 2013; Jang et al., 2019; Wen et al., 2019b; Jang et al., 2020; Sugihara et al., 2020; Vogel et al., 2021; Wang et al., 2021). Using BSA-seq, researchers can rapidly identify crucial genetic markers, facilitate targeted breeding efforts, and improve crop traits.

Based on NGS-based BSA-seq and genotyping with sequencing (GBS), Paudel et al. (2019) used three segregated F₂

populations (dotted black $\times$ green, dotted black $\times$ red, and dotted black $\times$ clump) to map the R, T1, W, and D loci on chromosomes 3, 5, 6, and 8, respectively. They developed allele-specific PCR (KASP™) assays and single-nucleotide polymorphism (SNP) markers for genotyping (Paudel et al., 2019). Additionally, Li et al. (2020) utilized an NGS-based genetic mapping approach with a light yellow and black seed coat RIL population to identify *Cla019481* on chromosome 3 as the gene responsible for black seed coat color (Li et al., 2020). In 2022, Maragal et al. employed a BC1F2 mapping population of BIL-53 (*Citrullus amarus*, red) $\times$ IHR-140-152 (*C. lanatus*, black) to identify three QTLs related to seed coat color: *q_SCC_3.1*, *q_SCC_5.1*, and *q_SCC_5.2*. Although several studies have been conducted, a genetic analysis using mapping populations generated by crossing parental lines harboring red and white seed coat colors in watermelons remains unexplored.

In this study, we aimed to identify candidate genomic regions in the red and white seed coats of watermelon using BSA-seq. The resulting delta SNP indices were analyzed by comparing sequences between two bulks from plants harboring red and white seed coats, and a SNP-based marker linked to seed coat color was developed by means of a high-resolution melting (HRM) analysis. This research provides insights into the genomic regions determining seed coat color in watermelon, facilitating marker-assisted seedling selection in watermelon breeding programs.

Materials and Methods

Plant material

The female and male parental lines used in this study were RG-21 (*C. lanatus*) (Fig. 1A) and Wr-609 (*C. amarus*) (Fig. 1B), which had white and red seed coats, respectively. Plant cultivation and seed production were conducted in a greenhouse using Partner Seeds (Gimje, South Korea). An F₁ hybrid was produced by crossing them in the summer of 2018 to create a mapped population. Thereafter, an F₂ generation was created by self-pollinating F₁ individuals in the summer of 2019. From the F₂ population, 219 individuals were grown for phenotypic surveys (Fig. 1C), and leaf samples were collected for DNA extraction.

Genetic analysis

We selected clear seed coat color phenotypes ($n = 80$) from the F₂ progeny for an inheritance analysis using a colorimetric measurement method. The goodness-of-fit between the expected and observed segregate ratios was analyzed using the chi-square test in R software (R Core Team, 2021).

Colorimetry measurements of seed coat colors

Intact seed coat colors were measured using a 3nh NR310 colorimeter (Shenzhen 3NH Tech. Co., Guangzhou, China). The light source was D65, and the excitation device was an LED blue light. Each measurement consisted of CIE L^* , a^* , and b^* , representing lightness and red/green and blue/yellow colors, respectively. Each seed was measured ten times, and the mean value was used in the data analysis. Based on these measurements, C^* (chroma: saturation of the color) and h_{ab} (hue angle: purity of the color) were calculated using the following equations:

$$C^* = (a^2 + b^2)^{1/2} \text{ and } h_{ab} = \frac{\arctan\left(\frac{b}{a}\right)}{2\pi} \times 360 \quad (1)$$

Principal component analysis (PCA) was utilized to assess the values of CIE L^* , C^* , and h_{ab} with the FactoMineR package (Lê et al., 2008), with visualization conducted using the ggplot2 package (Wickham, 2016) in R software ver. 4.1.2 (R Core Team, 2021).

DNA isolation

Leaf samples were collected for DNA isolation from individually tagged plants grown in a greenhouse (before they matured sufficiently). Genomic DNA was extracted using the GeneAll® Exgene™ Plant SV mini kit (GeneAll® Biotechnology, Seoul, Korea) in accordance with the manufacturer's instructions, with appropriate modifications. The quality of the extracted DNA was monitored using a Nanodrop spectrophotometer (DS-11, Denovix, Wilmington, DE, USA) and 1.2% agarose gel electrophoresis. The purified DNA samples were quantitatively analyzed using a Qubit 2.0 fluorometer (Invitrogen, Carlsbad, CA, USA) using Qubit™ dsDNA HS assay kits (Invitrogen), diluted to a working dilution of 20 ng/μL, and stored in a refrigerator at -20°C .

BSA-seq analysis

In this procedure, 30 plants with an extremely white seed coat (WSC) phenotype and 30 plants with an extremely red seed coat (RSC) phenotype were selected from the F_2 progeny for BSA, and 5 μg DNA was pooled from each individual to create WSC and RSC bulks. The two bulk and parental lines were sequenced at a sequencing depth of approximately 30× using a NovaSeq 6000 system (Illumina, San Diego, CA, USA). The resulting 'fastq' files were filtered using FastQC (Wingett and Andrews, 2018), and adapter sequences were eliminated using Trimmomatic v 0.39 (Bolger et al., 2014). High-quality paired reads obtained were aligned to the watermelon (97103) v2 genome (Guo et al., 2019) by means of Burrows – Wheeler alignment (BWA ver.0.7.17-r1188) with the mem option (Li and Durbin, 2009). Inaccurate paired reads were filtered, sorted, and deduplicated using "samtools" (Li et al., 2009) to produce BAM files. The variant call format (VCF) was generated via the "mpileup" command using BCF tools (Li, 2011). The VCF was filtered based on a sequencing depth of > 30 , a minor allele frequency of > 0.01 , and a missing sequence rate of 10%. The SNP index was estimated as the genomic difference between the WSC and RSC bulk and filtered by homo-SNPs in the WSC parental line of RG-21. Plots with options, including a 0.2 Mb window size and a 10 kb step size, were created using the QTL-seq program (Takagi et al., 2013). Positive and high SNP indices are significant genomic areas with seed coat color characteristics (Sugihara et al., 2020). Parental line sequences were used to identify hetero-SNPs and to construct PCR primers for an HRM investigation (Jang et al., 2020).

SNP markers for an HRM analysis

A genomic area with a high SNP index was determined to have a substantial QTL, and the Primer3 web tool (<https://bioinfo.ut.ee/primer3-0.4.0/>) was used to design the SNP-based primer sets (Rozen and Skaletsky, 2000). For the HRM analysis, PCR was performed using the LightCycler 96 platform (Roche, Mannheim, Germany). In a volume of 10

μL , 2 ng of dsDNA, $1\times$ LightCycler 480 High-Resolution Melting Master (Roche), $0.5\ \mu\text{M}$ of each primer, and 3 mM MgCl_2 were included in a reaction under the following conditions: 55 cycles at 94°C for 30 s, 60°C for 30 s, and 72°C for 30 s. The default HRM option was used for the final melting condition. HRM data were analyzed using the LightCycler[®] 96 SW 1.1 program (Roche Diagnostics, Mannheim, Germany). The amplified products were analyzed via electrophoresis on a 1.5% agarose gel containing $1\times$ TAE (Tris-acetate-EDTA) buffer at 150 V for 30 min, stained with ethidium bromide, and examined under UV light.

A total of 62 F_2 progeny previously used for bulk preparation with 34 HRM primer sets were examined to detect significant SNPs located at Chrs. 5 and 6 and to evaluate the significant genomic regions harboring a high ΔSNP index in the BSA-seq analysis. Next, 19 HRM primer sets were prepared and 136 additional F_2 progenies were analyzed to test for co-segregation between the tested markers and the seed coat color.

Results

Phenotypic segregation and inheritance of seed coat color

The overall phenotypes of the seed coat color were largely classified into white (with a tan tip), tan, and red in the family (Fig. 1). All seeds from the F_1 generation, in which color was absent from the parental lines, had tan seed coats (TSC) only (Fig. 1C and Table 1). The variability in the seed coat color of watermelons due to the interbreeding between RG-21 (*C. lanatus* with a white seed coat) and Wr-609 (*C. amarus* with a red seed coat) resulted in inconsistent seed maturity and reduced seed production. This motivated the selection of 80 individuals of F_2 progeny required for a meticulous analysis of color segregation, ensuring the representation of seed coat colors of the maternal, paternal, and F_1 generations. The seed

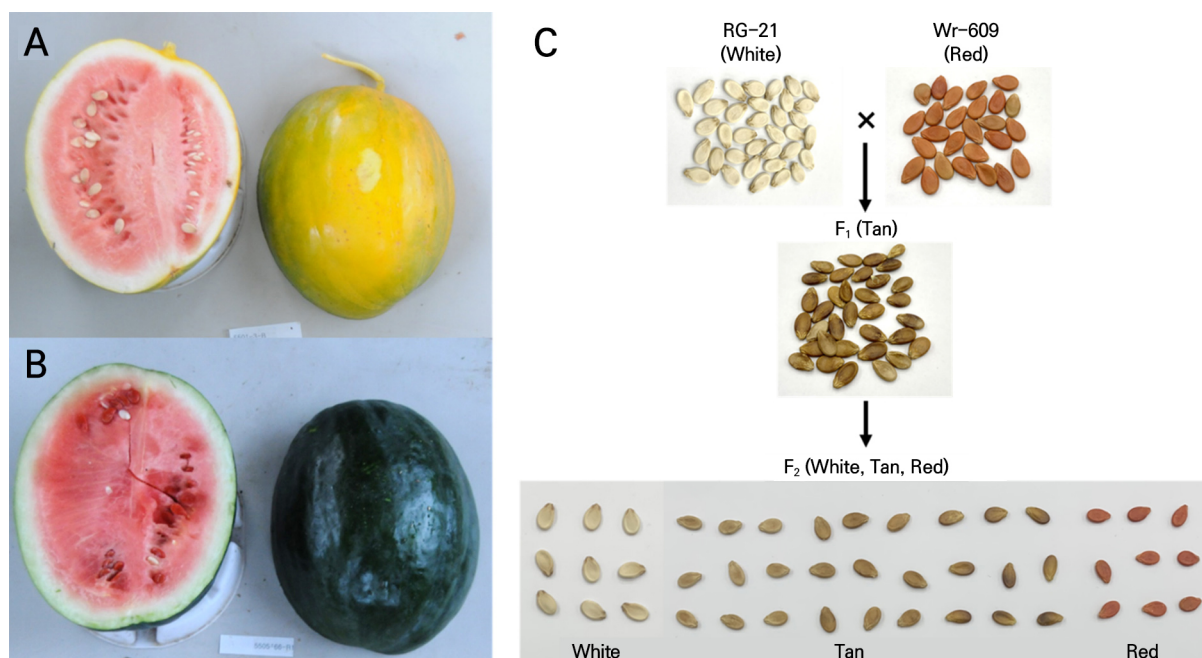


Fig. 1. Phenotype of parental lines and pedigree to produce the segregant population associated with seed coat color. (A) RG-21 (female, white seeds), (B) Wr-609 (male, red seeds); and C, pedigree from parent to F_2 generation.

Table 1. Phenotyping of seed coat color and chi-square test results of F₂ populations (RG-21 × Wr-609, n = 80)

Generation	Number of seed coat color			Expected	df	χ^2	p -value ^z
	White	Tan	Red				
P1 (RG-21)	30	0	0				
P2 (Wr-609)	0	0	30				
F1 (RG-21 × Wr-609)	0	30	0				
F2 (RG-21 × Wr-609)	14	48	18	4:9:3	2	2.6	0.2725

^z p -value presented from a goodness-of-fit test.

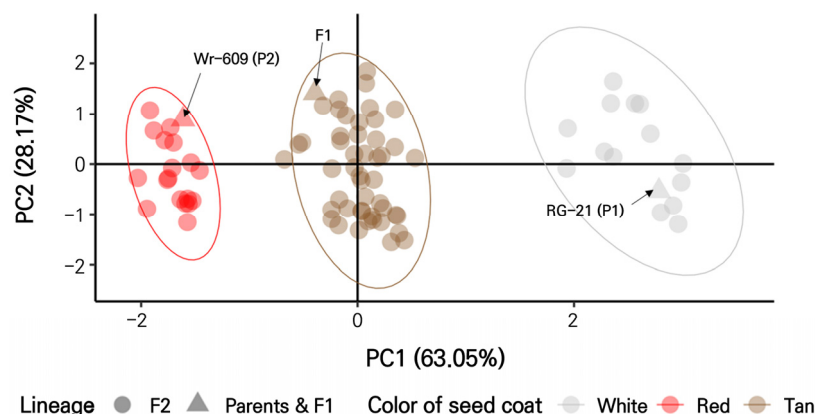


Fig. 2. Principal component analysis (PCA) analysis clustering of the F₂ phenotype based on seed coat color. The three main genotypic groups within the F₂ segregating population, as revealed by phenotyping, were as follows: the red seed coat group, phenotypically corresponding to the parental line Wr-609; the tan seed coat group, matching the F₁ phenotype; and the white seed coat group, aligned with the parental line RG-21. PCA analyses clustered these groups in accordance with the observed phenotypes in the F₂ individuals analyzed in this study.

coat colors were segregated into 14 white, 48 tan, and 18 red individuals. Based on the digitalized optical phenotyping results obtained from the F₂ progeny, we conducted a two-dimensional PCA to determine the variance in the segregated individuals (91.22%) explained by each principal component: PC1 (63.05%) and PC2 (28.17%) (Fig. 2). In the PCA visualization, the segregated F₂ individuals clustered into three groups and were separated with a difference in the measured CIE L^* , C^* , and h_{ab} variables between the groups. The first group exhibited a red seed coat, which was phenotypically assigned to the color of the parental line Wr-609. The second group had a tan-seed coat corresponding to the F₁ phenotype. The third group had a white seed coat that matched that of the parental line RG-21. PCA confirmed that these groups clustered in alignment with the phenotypes observed in the F₂ individuals analyzed in this study.

The segregation ratio for the F₂ progeny inferred according to the previously reported inheritance theory (Poole et al., 1941) for seed coat color is typically 9:3:3:1 for tan ($RrWw$), red ($rrWW$), white with a tan tip ($RRww$), and white with a pink tip ($rrww$). However, in our study, we observed white seed coats without colored tips, leading us to hypothesize a modified segregation ratio of 9:3:4. To test this hypothesis, we conducted a chi-square analysis ($\alpha = 0.05$) using the results from PCA as the basis for our assumption. The result indicated that the hypothesis of 9:3:4 segregation is not rejected ($\chi^2 = 2.6$, p -value = 0.2725; Table 1), suggesting that seed coat color is controlled by two genes with negative epistasis.

Whole-genome resequencing and mapping

We aimed to understand the genetic inheritance through further research, in this case a BSA-seq analysis. Table 2 displays the whole-genome resequencing and paired-mapping rates. When matched with the reference watermelon genome, the total reads produced for the WSC and RSC bulks exceeded 80 million, and the appropriate paired-mapping rate was greater than 86% (Guo et al., 2019). When comparing BSA bulks at the genome level, the average depth for all covered regions containing high-quality reads was greater than 34, which was considered adequate (Abe et al., 2012; Jang et al., 2020).

Scanning candidate genomic regions for seed coat color

The final 380,726 genomic variations between the WSC and RSC bulks from the F₂ progeny were filtered using RG-21 (WSC) sequences as controls to determine the SNP index (Takagi et al., 2013; Itoh et al., 2019). The major loci on Chr. 5 (3.6–7.7 Mb, $p < 0.05$) and 6 (5.1–9.7 Mb, $p < 0.05$) were discovered between Chr. 1 and 11 of the watermelon genome when we simulated the average SNP index in a 2.0 Mb interval using a 10 kb increment sliding window (Fig. 3 and Suppl. Fig. S1).

Localization of the WSC and RSC candidate genomic regions according to SNP markers and physical mapping

In the initial BSA-seq analysis, we designed 34 HRM primer sets targeting regions with distinct variants from the parental line, ranging from 3.6 Mb to 7.7 Mb on Chr. 5 and from 5.1 Mb to 9.7 Mb on Chr. 6. We then tested these primer

Table 2. Whole-genome re-sequencing and mapping statistics of parental lines and the two bulks of white seed and red plants

Genotype	Number of samples	Number of generated reads (bp)	Total bases (Mb)	Properly paired mapping rate (%)	Mean depth (X) for all covered regions
RG-21	1	85,122,026	11,253	91.54	25.28
Wr-609	1	82,354,390	10,179	90.15	24.88
White seed bulk	30	83,236,492	10,305	86.45	24.90
Red seed bulk	30	82,484,624	9,802	86.22	23.64

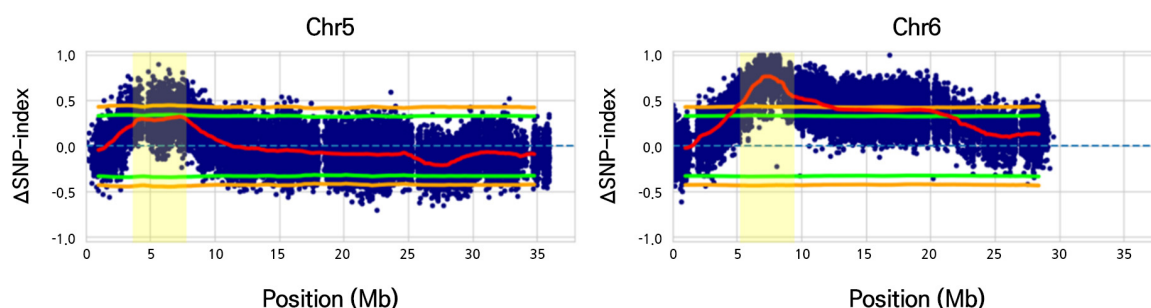
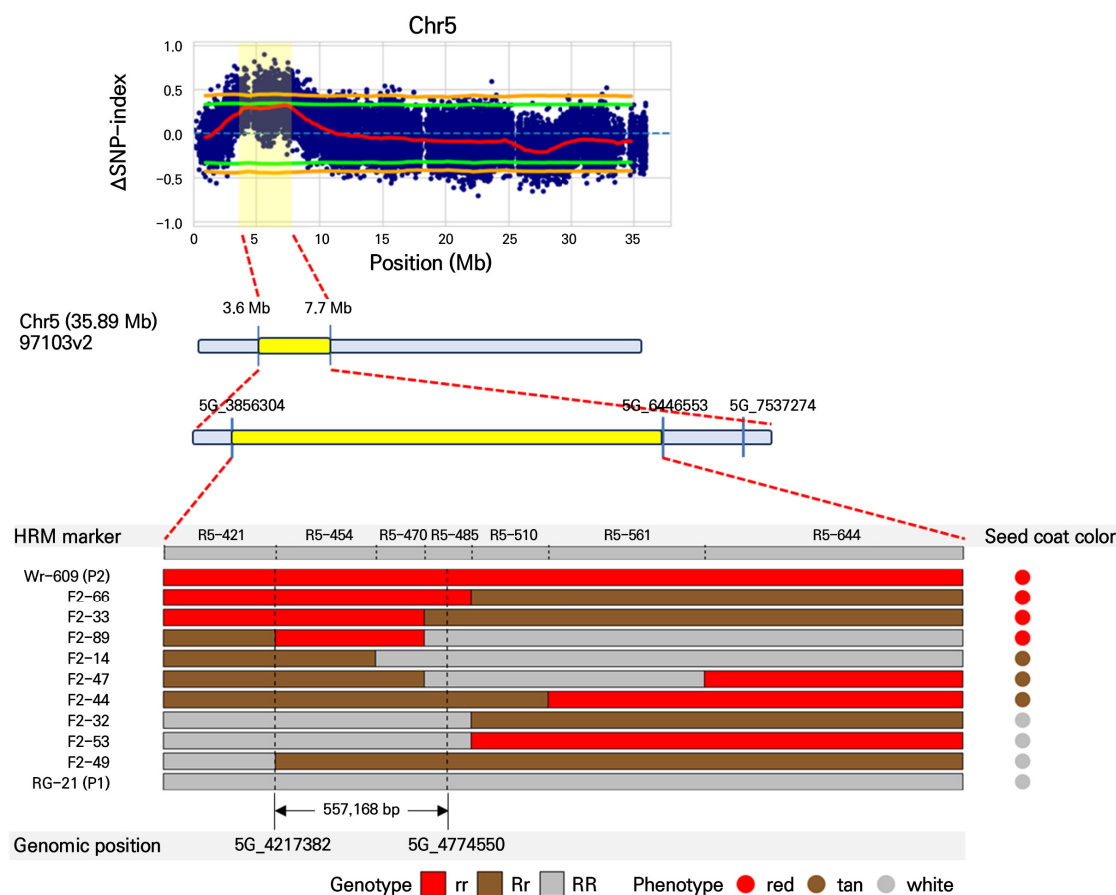


Fig. 3. Mapping of candidate genomic regions for seed coat color on chromosomes 5 and 6. The Δ SNP index represents the difference between the SNP index of the white seed coat bulk and the red seed coat bulk. Significant candidate genomic regions are enclosed by dotted boxes and are highlighted in yellow. The red line represents the Δ SNP index; the orange line indicates $p < 0.01$; the green line denotes $p < 0.05$.

Table 3. Validation of the high-resolution melting (HRM) analysis of 19 single-nucleotide polymorphism (SNP) markers

Chromosome	Location (bp)	SNP	Primer ^z	Co-segregation ratio (%)
5	4,218,433	C > A	R5-421	82
	4,542,806	G > A	R5-454	92
	4,700,480	T > C	R5-470	92
	4,852,402	G > T	R5-485	85
	5,101,217	A > G	R5-510	67
	5,610,509	A > G	R5-561	58
	6,446,553	G > A	R5-644	58
6	6,841,088	C > T	W6-684	89
	6,928,966	C > T	W6-692	95
	7,008,987	G > T	W6-700	100
	7,121,807	C > A	W6-712	100
	7,403,446	C > T	W6-740	92
	7,566,534	C > T	W6-756	85
	7,826,437	T > A	W6-782	85

^zPrimer sequences are provided in [Suppl. Table S1](#).**Fig. 4.** Physical fine mapping of the genomic region associated with red seed coat color on chromosome 5 in watermelon. The Δ SNP index plot shows the difference between the SNP index of the white seed coat bulk and the red seed coat bulk along chromosome 5. The significant candidate regions are highlighted in yellow. The physical fine mapping narrows down the candidate region further to a specific interval between 3.6 Mb and 7.7 Mb, with a delimited region of 557,168 bp. Genotyping of the F₂ progeny and parental lines revealed a clear segregation pattern corresponding to the phenotypic expression of seed coat color. Red segments indicate the red seed coat phenotype (genotype rr), and tan and white segments correspond to the tan (Rr) and white (RR) phenotypes, respectively.

sets on 60 F₂ progeny individuals, which were used for bulk preparation in the BSA-seq analysis, to focus on the candidate regions of Chr. 5 and Chr. 6. To narrow the range of the two candidate loci further, we analyzed nine additional primer sets on Chr. 5 and ten additional primer sets on Chr. 6 in 106 F₂ progeny. In the RSC case, the highest co-segregation ratio of SNP markers (92%) ranged from 4,542,806 to 4,700,480 bp on Chr. 5 (Table 3). In the WSC trait, we found that the SNP marker located at 7,008,987 to 7,121,807 on Chr. 6 matched the WSC phenotype perfectly (Table 3).

The physical fine-mapping results identified candidate genomic regions associated with seed coat color in watermelon. In Fig. 4, fine mapping pinpoints a specific interval between 3.6 Mb and 7.7 Mb on Chr. 5, narrowing down to a region of 557,168 bp associated with red seed coat color. Additionally, in Fig. 5, the region was delimited to an interval between 5.1 Mb and 9.7 Mb on Chr. 6, narrowing to a region of 207,137 bp associated with white seed coat color. Genotyping of the F₂ progeny and parental lines revealed a clear segregation pattern that aligned with the phenotypic expression of seed coat color. The red seed coat (genotype *rr*) and white seed coat (genotype *ww*) phenotypes were predominantly linked to the identified candidate regions, as indicated by the consistent presence of red and white genotype segments in individuals expressing red and white seed coat colors (Figs. 4 and 5).

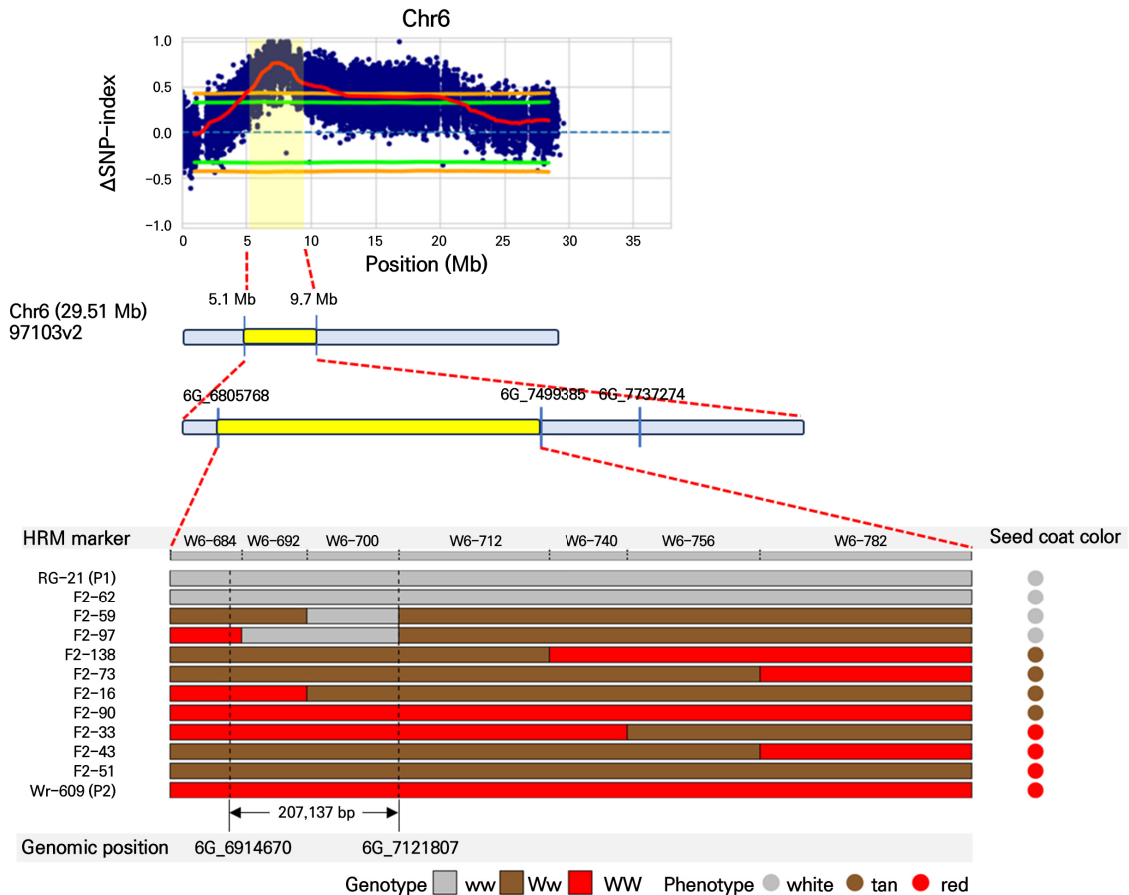


Fig. 5. Physical fine mapping of the genomic region associated with white seed coat color on chromosome 6 in watermelon. The Δ SNP index plot shows the difference between the SNP index of the white seed coat bulk and the red seed coat bulk along chromosome 6. The significant candidate regions are highlighted in yellow. The physical fine mapping narrows down the candidate region further to a specific interval between 5.1 Mb and 9.7 Mb, with a delimited region of 207,137 bp. Genotyping of the F₂ progeny and parental lines revealed a clear segregation pattern corresponding to the phenotypic expression of seed coat color. White segments indicate the white seed coat phenotype (genotype *ww*), tan segments correspond to the tan phenotype (*Ww*), and red segments represent the red phenotype (*WW*).

Discussion

Several genes with intricate genetic connections are responsible for the genetic determination of seed coat color (Poole, 1941). In the first investigation into the inheritance of seed coat color (Kanda, 1931), flat black seed coat color was found to be monogenically dominant over dotted black seed coat color by crossing flat black watermelon seeds with dotted black watermelon seeds. Later, McKay (1936) developed tan red and green red hybrids and demonstrated that the former phenotype is monogenically dominant over the latter in each cross (McKay, 1936). The dotted black seed coat is monogenically dominant over the clump seed coat according to Weetman (1937), who also demonstrated that various combinations of the two genes create clump and tan seed coat colors (Weetman, 1937).

A genetic model consisting of the four genes, *R*, *T*, *W*, and *D*, was developed in the 1940s to explain the inheritance of seed coat color in watermelons (Poole, 1941). Poole et al. (1941) created 40 distinct segregating populations and suggested a four-gene hypothesis for determining watermelon seed coat color. The seed coat colors produced by various combinations of the *R*, *T*, and *W* genes with the modifier gene *D* are flat black (*RTWD*), dotted black (*RTWd*), green (*rTW*), tan (*RtW*), clump (*RTw*), red (*rtW*), white-tan tip (*Rtw*), and white-pink tip (*rtw*) (Paudel et al., 2019). Additionally, Chrs. 3, 5, 6, and 8 were assigned to the *R*, *T*, *W*, and *D* genes, respectively. The seed color markers UGA3 5820134, UGA5 4591722, UGA6 7076766, and UGA8 22729513 were also created. In a recent study, the watermelon seed coat was bright yellow or black due to premature termination of protein translation due to a genetic mutation in the 70.2 kb region of Chr. 3 (Li et al., 2020).

In this study, the seed coat color of all F_1 hybrids from the cross between RG-21 (WSC) and Wr-609 (RSC) was tan. According to the seed coat color theory for watermelon (Poole et al., 1941; Paudel et al., 2019), the expected genotype for F_1 hybrids is *RrttWwdd*, derived from parental lines RG-21 (*RRttwwdd*, WSC) and Wr-609 (*rrttWWdd*, RSC). We did not consider the *T* and *D* loci in our analysis, as phenotypes such as flat black (*RRTTWWDD*), dotted black (*RRTTWWdd*), or clump (*RRTTww*) were absent. Consequently, the expected segregation ratio in the F_2 generation from a dihybrid cross of F_1 (*RrWw*) should typically follow a 9:3:3:1 pattern for tan, red, white with a tan tip, and white with a pink tip. However, our study found that the phenotypes white with a tan tip (*RRww*) and white with a pink tip (*rrww*) were indistinguishable, with only white with a white tip being observed. We hypothesized that negative epistasis between the *R* and *W* loci could alter the expected ratio to 9:3:4 (tan, red, white). The BSA-seq analysis suggested that each of these loci (*R* and *W*) is located on different chromosomes, lending support to this hypothesis. Upon a closer examination, we found that the genomic regions closely linked to the WSC phenotype were located at 6,914,670 and 7,121,807 bp on Chr. 6. These SNP markers for WSC were closely located in the genomic region controlling the *W* gene according to a previous report (Paudel et al., 2019). Hence, red in the RSC was closely linked to 4,217,382–4,774,550 bp on Chr. 5, to which the *T^l* locus for tan was assigned on Chr. 5 (Paudel et al., 2019). Although no variation in the tan phenotype in F_1 seeds was observed, the tan phenotype of the F_2 progeny varied, i.e., was not fixed. These results imply that the RSC gene located on Chr. 5 could regulate or interact with the *T^l* locus reported by Paudel et al. (2019). In previous studies, no specific molecular evidence supporting the red color or the genes involved in the red color could be found. Therefore, this study provides a basis for understanding the RSC traits of watermelons and for breeding new watermelon varieties.

Literature Cited

- Abe A, Kosugi S, Yoshida K, Natsume S, Takagi H, Kanzaki H, Matsumura H, Yoshida K, Mitsuoka C, et al. (2012) Genome sequencing reveals agronomically important loci in rice using MutMap. *Nat Biotechnol* 30:174. doi:10.1038/nbt.2095
- Baskin J, Baskin C, Li X (2000) Taxonomy, anatomy and evolution of physical dormancy in seeds. *Plant Species Biol* 15:139-152. doi:10.146/j.1442-1984.2000.00034.x
- Bolger AM, Lohse M, Usadel B (2014) Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinfo* 30:2114-2120. doi:10.1093/bioinformatics/btu170
- Dong W, Wu D, Li G, Wu D, Wang Z (2018) Next-generation sequencing from bulked segregant analysis identifies a dwarfism gene in watermelon. *Sci Rep* 8:2908. doi:10.1038/s41598-018-21293-1
- El-Bramawy ME-HS, El-Hendawy S, Shaban WA (2008) Assessing the suitability of morphological and phenological traits to screen sesame genotypes for fusarium wilt and charcoal rot disease resistance. *J Plant Prot Res* 48:397-410. doi:10.2478/v10045-008-0049-y
- Guner N, Wehner TC (2004) The genes of watermelon. *Hort Sci* 39:1175-1182. doi:10.21273/HORTSCI.39.6.1175
- Guo S, Zhang J, Sun H, Salse J, Lucas WJ, Zhang H, Zheng Y, Mao L, Ren Y, Wang Z (2013) The draft genome of watermelon (*Citrullus lanatus*) and resequencing of 20 diverse accessions. *Nat Genet* 45:51-58. doi:10.1038/ng.2470
- Guo S, Zhao S, Sun H, Wang X, Wu S, Lin T, Ren Y, Gao L, Deng Y, Zhang J (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
- Gusmini G, Wehner T, Jarret R (2004) Inheritance of egusi seed type in watermelon. *J Hered* 95:268-270. doi:10.1093/jhered/esh031
- Itoh N, Segawa T, Tamiru M, Abe A, Sakamoto S, Uemura A, Oikawa K, Kutsuzawa H, Koga H, et al. (2019) Next-generation sequencing-based bulked segregant analysis for QTL mapping in the heterozygous species *Brassica rapa*. *Theor Appl Genet* 132:2913-2925. doi:10.1007/s00122-019-03396-z
- Jang YJ, Seo M, Hersh CP, Rhee SJ, Kim Y, Lee GP (2019) An evolutionarily conserved non-synonymous SNP in a leucine-rich repeat domain determines anthracnose resistance in watermelon. *Theor Appl Genet* 132:473-488. doi:10.1007/s00122-018-3235-y
- Jang YJ, Yun HS, Rhee S-J, Seo M, Kim Y, Lee GP (2020) Exploring molecular markers and candidate genes responsible for watermelon dwarfism. *Hortic Environ Biotech* 61:173-182. doi:10.1007/s13580-020-00229-7
- Kanda T (1931) The inheritance of seed-coat coloring in the watermelon. *Japanese J Genet* 7:30-48. doi:10.1266/jjg.y.30
- Lê S, Josse J, Husson F (2008) FactoMineR: An R package for multivariate analysis. *J Stat Soft Vol* 25. doi:10.18637/jss.v025.i01
- Li B, Lu X, Gebremeskel H, Zhao S, He N, Yuan P, Gong C, Mohammed U, Liu W (2020) Genetic mapping and discovery of the candidate gene for black seed coat color in watermelon (*Citrullus lanatus*). *Front Plant Sci* 10:1689. doi:10.3389/fpls.2019.01689
- Li H (2011) A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. *Bioinfo* 27:2987-2993. doi:10.1093/bioinformatics/btr509
- Li H, Durbin R (2009) Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinfo* 25:1754-1760. doi:10.1093/bioinformatics/btp324
- Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R (2009) The sequence alignment/map format and SAMtools. *Bioinfo* 25:2078-2079. doi:10.1093/bioinformatics/btp352
- Mavi K (2010) The relationship between seed coat color and seed quality in watermelon Crimson sweet. *Hortic Sci* 37:62-69. doi:10.1722/1/53/2009-HORTSCI
- McKay J (1936) Factor interaction in *Citrullus*: Seed-coat color, fruit shape and markings show evidence of mendelian inheritance in watermelon crosses. *J Hered* 27:110-112. doi:10.1093/oxfordjournals.jhered.a104182
- Paudel L, Clevenger J, McGregor C (2019) Chromosomal locations and interactions of four loci associated with seed coat color in watermelon. *Front Plant Sci* 10:460813. doi:10.3389/fpls.2019.00788
- Poole C (1944) Genetics of cultivated cucurbits. *J Heredity* 35:122-128. doi:10.1093/oxfordjournals.jhered.a105364
- Poole CF (1941) VOL. 63 Washington, Dc, October 15, 1941 No. 8 Inheritance of seed characters in water-melon. *J Agric Res* 63:433
- Poole CF, Grimbail PC, Porter DR (1941) Inheritance of seed characters in watermelon. *J Agric Res* 63:433-456
- Powell A (1989) The importance of genetically determined seed coat characteristics to seed quality in grain legumes. *Ann Bot* 63:169-175. doi:10.1093/oxfordjournals.aob.a087720
- R Core Team (2021) R: A language and environment for statistical computing. <https://www.r-project.org>. Accessed 20 March 2024
- Rimando AM, Perkins-Veazie PM (2005) Determination of citrulline in watermelon rind. *J Chromatogr A* 1078:196-200. doi:10.1016/j.chroma.2005.05.009
- Rozen S, Skaletsky H (2000) Primer3 on the WWW for general users and for biologist programmers. *Methods Mol Biol* 365:386. doi:10.1385/1-59259-192-2:365
- Ruangrak E, Su X, Huang Z, Wang X, Guo Y, Du Y, Gao J (2018) Fine mapping of a major QTL controlling early flowering in tomato using QTL-seq. *Can. J Plant Sci* 98:672-682. doi:10.1139/cjps-2016-0398
- Sugihara Y, Young L, Yaegashi H, Natsume S, Shea DJ, Takagi H, Booker H, Innan H, Terauchi R, et al. (2020) High-performance pipeline for MutMap and QTL-seq. *PeerJ Vol* 10. doi:10.1101/2020.06.28.176586
- Takagi H, Abe A, Yoshida K, Kosugi S, Natsume S, Mitsuoka C, Uemura A, Utsushi H, Tamiru M, et al. (2013) QTL-seq: rapid mapping of quantitative trait loci in rice by whole genome resequencing of DNA from two bulked populations. *Plant J* 74:174-183. doi:10.1111

/tpj.12105

- Vogel G, LaPlant KE, Mazourek M, Gore MA, Smart CD** (2021) A combined BSA-Seq and linkage mapping approach identifies genomic regions associated with *Phytophthora* root and crown rot resistance in squash. *Theor Appl Gene* 134:1015-1031. doi:10.1007/s00122-020-03747-1
- Wang Z, Yu A, Li F, Xu W, Han B, Cheng X, Liu A** (2021) Bulk segregant analysis reveals candidate genes responsible for dwarf formation in woody oilseed crop castor bean. *Sci Rep* 11:1-15. doi:10.1038/s41598-021-85644-1
- Weetman LM** (1937) Inheritance and correlation of shape, size and color in the watermelon, *Citrullus vulgaris* Schrad. *Iowa Agriculture and Home Economics Experiment Station Research Bulletin* 20:1
- Wehner TC** (2008) Watermelon. In *Vegetables I*. Springer, pp 381-418. doi:10.1007/978-0-387-30443-4_12
- Wen C, Zhang J, Feng Y, Duan Y, Ma H, Zhang H** (2020) Purification and identification of novel antioxidant peptides from watermelon seed protein hydrolysates and their cytoprotective effects on H₂O₂-induced oxidative stress. *Food Chem* 327:127059. doi:10.1016/j.foodchem.2020.127059
- Wen C, Zhang J, Zhang H, Duan Y, Ma H** (2019a) Effects of divergent ultrasound pretreatment on the structure of watermelon seed protein and the antioxidant activity of its hydrolysates. *Food Chem* 299:125165. doi:10.1016/j.foodchem.2019.125165
- Wen J, Jiang F, Weng Y, Sun M, Shi X, Zhou Y, Yu L, Wu Z** (2019b) Identification of heat-tolerance QTLs and high-temperature stress-responsive genes through conventional QTL mapping, QTL-seq and RNA-seq in tomato. *BMC Plant Biol* 19:1-17. doi:10.1186/s12870-019-2008-3
- West S, Harris H** (1963) Seedcoat Colors Associated with Physiological Changes in Alfalfa and Crimson and White Clovers 1. *Crop Sci* 3:190-193. doi:10.2135/cropsci1963.0011183X000300030005x
- Wickham H** (2016) ggplot2: Elegant graphics for data analysis. Springer-Verlag, New York doi:10.1007/978-3-319-24277-4_9
- Wingett SW, Andrews S** (2018) FastQ Screen: A tool for multi-genome mapping and quality control. *F1000Res* 7:1338. doi:10.12688/f1000research.15931.1
- Wu S, Wang X, Reddy U, Sun H, Bao K, Gao L, Mao L, Patel T, Ortiz C, Abburi VL** (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