Article

Assessment of the Genetic Diversity of the Breeding Lines and a Genome Wide Association Study of Three Horticultural Traits Using Worldwide Cucumber (*Cucumis* spp.) Germplasm Collection

Hea-Young Lee [1], Jeong-Gu Kim [2], Byoung-Cheorl Kang [3] and Kihwan Song [1,*]

[1] Department of Bioresources Engineering, Sejong University, Seoul 05006, Korea; hylee@sejong.ac.kr
[2] National Academy of Agricultural Science, Rural Development Administration, Jeonju 54874, Korea; jkim5aug@korea.kr
[3] Department of Plant Science, Plant Genomics and Breeding Institute, College of Agriculture and Life Science, Seoul National University, Seoul 08826, Korea; bk54@snu.ac.kr
[*] Correspondence: khsong@sejong.ac.kr; Tel.: +82-2-3408-2905

check for updates

Abstract: Cucumbers are an important economic vegetable crop that is used for fresh or processing purposes worldwide. In this study, we used 264 accessions that consisted of world-wide wild germplasms and advanced breeding lines in order to understand the genetic diversity and the genetic correlation among the germplasm collection. A genotyping-by-sequencing (GBS) approach was applied to obtain dense genome-wide markers coverage (>12,082 SNPs) to construct a high-density haplotype map. Various population stratification methods were performed, and three subgroups were divided based on the genetic diversity, which reflected their geographic regions. According to the phylogenetic analysis, the breeding lines were separated from wild germplasms, and the two distinct groups were divided within the breeding lines. One of the groups mainly consisted of East-Asian varieties, which showed the unique homogenous genotype patterns. Using this germplasm collection, three important horticultural traits of cucumbers—powdery mildew resistance, spine color, and fruit stalk-end color—were evaluated and used to conduct the genome-wide association study (GWAS). All of the significant SNPs and two novel candidate genes (*Csa5G453160* and *Csa5G471070*) for the powdery mildew were identified in chromosome 5 from the natural population, which is where reported major QTLs from various bi-parental population are located. Furthermore, two candidate genes, *Csa1G006300* and *Csa3G824850*, and four candidate genes, *Csa2G368270*, *Csa3G236570*, *Csa5G175680*, and *Csa6G448170*, were identified for the spine color and the fruit stalk-end color, respectively. These results are expected to be helpful to develop molecular markers of the horticultural traits in cucumbers.

Keywords: cucumber; genotyping-by-sequencing (GBS); germplasm diversity; breeding lines; genome-wide association study (GWAS)

1. Introduction

Cucumbers (*Cucummis* spp.) are an economically important vegetable crop worldwide that are used for salads, pickles, and fresh consumption [1]. They are rich in dietary fiber; vitamins, which include vitamins B, C, and K; antioxidants; and minerals, such as potassium and magnesium [2]. The origin of the cucumber is known to be India, which is where the wild ancestor *C. sativus* var. *hardwickii* round and bitter fruit grows [1,3,4]. During the last 3000 years, various varieties adapted to the local environmental conditions and the food culture have been differentiated [4,5]. Modern cucumber

cultivars show diverse variation in fruit characteristics, such as shapes, colors, sizes, skin textures, and bitterness depending on the cultivation area and the consumption methods [1,6]. The diversity could be found mainly in fruit shape, which includes pickling cucumbers in America; long green slicers such as the Long Dutch type in North Europe; the mini cucumber, which is the Beith Alpha type in South Europe; and the long green cucumbers of various color shades in many Asian countries. Even though there exist numerous variations in the fruit morphology, the current local cucumber varieties seem to have a narrow genetic diversity as the regional-specific domestication and breeding have progressed. In fact, in order to develop stress tolerance varieties in response to rapid climate change, it is necessary to secure and preserve various wild germplasms that possess potentially useful traits.

Therefore, in order to overcome these genetic limitations, a large number of germplasms and resources have been collected, preserved, and managed in many ways [7]. One of the effective managements of germplasms is construction of the core collection. Cucumber core collections have been constructed and evaluated to preserve the genetic diversity with an effective minimum population [8,9]. According to these studies, cucumber germplasms could be divided into three or four groups based on the population structure that reflects their geographical isolation [8–10]. However, they focused only on broad genetic diversity and the population structure, not on details. It is important to understand the genetic diversity of the local germplasms, because the cucumber varieties have been developed with strong regional characteristics. Furthermore, understanding the genetic diversity could provide a foundation for a breeding strategy to develop regionally adapted varieties in order to cope with the global climate change.

The cucumber genome assemblies have been released and updated in three types of cucumbers—the Chinese long (9930), the American pickling cucumber (Gy14), and the North European inbreeding (B10) line from Borszczagowski [7,11–13]. Using these assemblies, numerous SNP arrays and markers were developed and applied to speed up the genetic studies in cucumbers [14]. Based on a large number of SNPs, the genetic association studies were conducted on general economical traits, such as resistance to nematodes, downy mildew, and powdery mildew [15–17]. According to the inheritance studies, it is known that the PM resistance is controlled by multiple recessive genes [16,18–21]. The QTLs for the PM have been identified using several inbred lines and the RILs [16,18–20,22]. Among the QTLs, the QTL *pm 1.1* and the *pm 5.2* are effective in leaf and hypocotyl resistance [20]. Some candidate genes of powdery mildew resistance were suggested, which included the MLO-like gene [21,23], *Csa1M064780*, *Csa1M064790* [24], the *CsSGR* gene [25], the cyclin-like (cyclin L) gene, genes encoding the F-box protein, kinesin protein, calcium transporting ATPase, Ras-related protein, and lysine-specific histone demethylase 1 (LSD1) [15]. Interestingly, the powdery mildew resistance is reported to be closely linked to the downy mildew resistance [16,22]. Since all of the QTL and the associated genes were discovered from various bi-parental populations, confirmation within the natural population is highly needed to exclude the population-specific bias results.

QTLs, genes, and several correlated loci of horticultural traits, such as the flowering time, fruit sizes, fruit shapes, fruit firmness, and the pericarp color have been also identified [26–28]. The color of spine and the fruit's stalk end in cucumbers are important consumer traits that determine the fruit quality and freshness. Despite the importance, studies on the fruit's stalk end color are very rare. The spine color has been reported to be regulated by a single dominant *B* gene [29,30], or two genes, which are *B* and *B2* identified from cultivated cucumbers or *B3* and *B4* identified from crossing with wild cucumbers [31,32]. In a recent study, *R2R3-MYB* is considered as a candidate *B* gene [29], and the *R2R3-MYB*, which is called *HEUKCHEEM*, is involved in controlling the black spines of cultivated cucumbers [30]. According to the latest study, two different mutations—the insertion of a 6994 bp transposable element and one bp non-synonymous SNP in *HEUKCHEEM*—caused the white spines [30].

In this study, we studied the genetic diversity and the population structure of the 264 cucumber accessions that consist of wild germplasms and advanced breeding lines, and we focused on the comparison of the genetic relationship between the breeding lines and the wild germplasms.

Furthermore, high quality genome-wide SNPs via genotyping-by-sequencing (GBS) were applied to identify the significant genetic regions correlated with the three horticultural traits—the PM, the fruit's stalk-end color, and the spine color—of cucumbers through genome-wide association studies (GWAS).

2. Materials and Methods

2.1. Plant Materials

A cucumber germplasm collection that consisted of 264 accessions, which included the local wild varieties and the breeding lines, was assessed in this study (Table S1). This population consisted of four species that were collected from 37 countries. Among the *Cucumis* species, most of the accessions belong to *C. sativus* (257 accessions), and the other three species *C. anguria*, *C. myriocarpus*, and *C. zeyheri* accounted for the 3, 3, and 1 accession, respectively. Among the accessions, 100 accessions are the advanced breeding lines that were developed for commercial breeding, and the other 164 accessions are the wild germplasms that are systematically sampled from the USDA National Plant Germplasm System (NPGS) based on the diversity.

2.2. Phenotypic Evaluation and Correlation Analysis

Over a two-year period (2018–2019), the cucumber germplasm was planted in two seasons, spring and autumn, in plastic houses at Anseong-si, Korea. Two replications with three plants of each accession were evaluated throughout the growing period. For the GWAS, three important horticultural traits, the powdery mildew resistance, the fruit's stalk-end color, and the spine color were evaluated in four different environments.

The resistance to powdery mildew at the seedling stage was evaluated in addition to the field resistance. The seedlings with fully expanded cotyledons were inoculated three times during three-day intervals, and the conidial suspensions collected from the diseased plants. The density of the spore suspension was adjusted to approximately 1×10^5 sporangia per ml prior to the inoculation. After the inoculation, the seedlings were kept in humid plastic tunnels for a day. The resistance index was scored at 2~3 weeks after the inoculation based on the resistance scale of 1–9, where 1: wilting and death of the whole plant (completely susceptible); 3: infection of the whole leaf (susceptible); 5: infection less than 30% of the leaf area (intermediate resistant); 7: several infected spots on the leaf surface (resistant); and 9: no visible symptoms (strong resistant). The fruit's stalk-end color (shoulder color) was scored as 1 to 3 (1 = short green stalk-end, 2 = long green stalk-end, and 3 = whole green body) (Figure S1). The fruit's spine color was scored from 1 to 3 (1 = white, 2 = brown, and 3 = black).

2.3. Genomic DNA Extraction and GBS

Genomic DNA was extracted from the samples using the CTAB method [33] and diluted to 80 ng/μL in distilled water. The GBS libraries were constructed via double digest restriction enzymes (EcoRI/MseI), and a compatible set of 96 barcodes was used as previously described [34]. The digested DNA was ligated to adapters and amplified with TA primers. The libraries were pooled in three tubes. The contents of the tubes were sequenced in separate lanes using the HiSeq 2000 platform (Illumina, San Diego, CA, USA) at Macrogen (Seoul, Korea).

2.4. Reference-Based SNP Calling and Construction of the SNP Set

Raw 101-bp reads from the libraries were trimmed to a minimum length of 80 bp and filtered to a quality score > Q30. The filtered reads were aligned to the *Cucumis sativus* L. var. *sativus* cv. 9930 (Chinese Long) as a reference genome sequence [11] using the Burrows–Wheeler Aligner program v.0.7.12 [35]. For the SNP calling and filtering, the GATK Unified Genotyper v.3.3 was used [36]. Raw SNPs were filtered to remove the mono and tri-allelic SNP types. The final SNP set was constructed after filtering, which followed the conditions that are listed next. Minor allele frequency > 0.05, SNP coverage > 0.6, and inbreeding coefficient (IF) > 0.8.

2.5. Population Structure Analysis

To analyze the population structure, we used a model-based genetic clustering algorithm [37] that was implemented in the STRUCTURE program version 2.3.4 [38]. The number of subpopulations (ΔK) was determined using the ad hoc statistical method, which is based on the rate change in the log probability of the data between the successive K values [39]. Ten independent runs for the K values that ranged from 1 to 10 were performed with a burn-in length of 50,000 followed by 1,000,000 iterations.

2.6. Phylogenetic and the Principal Coordinate Analyses

The genetic relationships among the inferred subpopulations were investigated based on phylogenetic and principal coordinate analyses. The phylogenetic trees were produced using the weighted neighbor-joining clustering method based on the dissimilarity matrix with software DARwin 6.0.9 [40]. The principal component analysis (PCA) was performed using the "pca Methods" library in R software [41]. The values of each PC were used as variables in the GWAS.
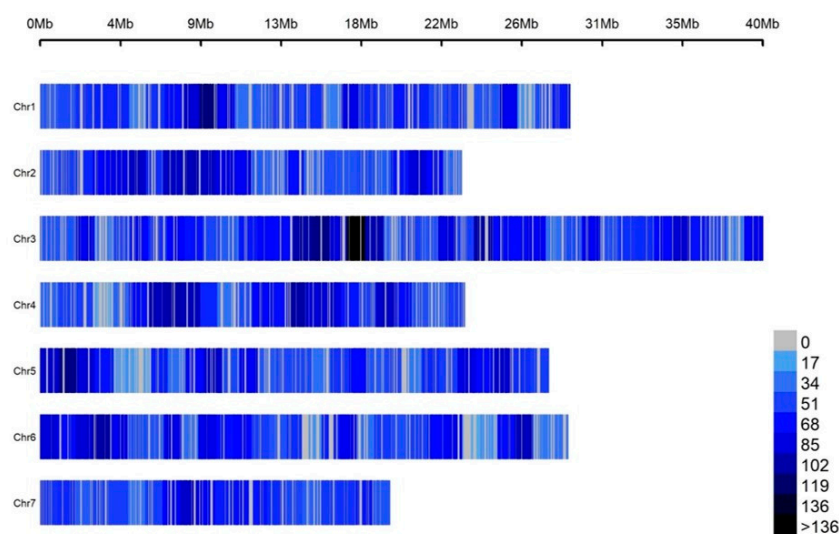
2.7. Genome-wide Association Study (GWAS) and Candidate Gene Identification

The GWAS based on the efficient mixed model association (EMMA) was conducted using the R package of Genomic Association and Prediction Integrated Tool (GAPIT) [42] with the default settings. The significance threshold $-\log_{10}$ *p*-value of the GWAS was determined using the Bonferroni correction (FDR *p*-value < 0.05) [43] based on the number of independent SNPs in a population. Version 2 of the Chinese Long gff3 file (cucurbitgenomics.org/pub/cucurbit/genome/cucumber/Chinese_long/v2) was used to identify the nearby genes of each significant SNP from the GWAS results. The candidate genes were selected by comparing them with the previously reported genes or by assessing the 1 Mbp regions around the significant SNP.
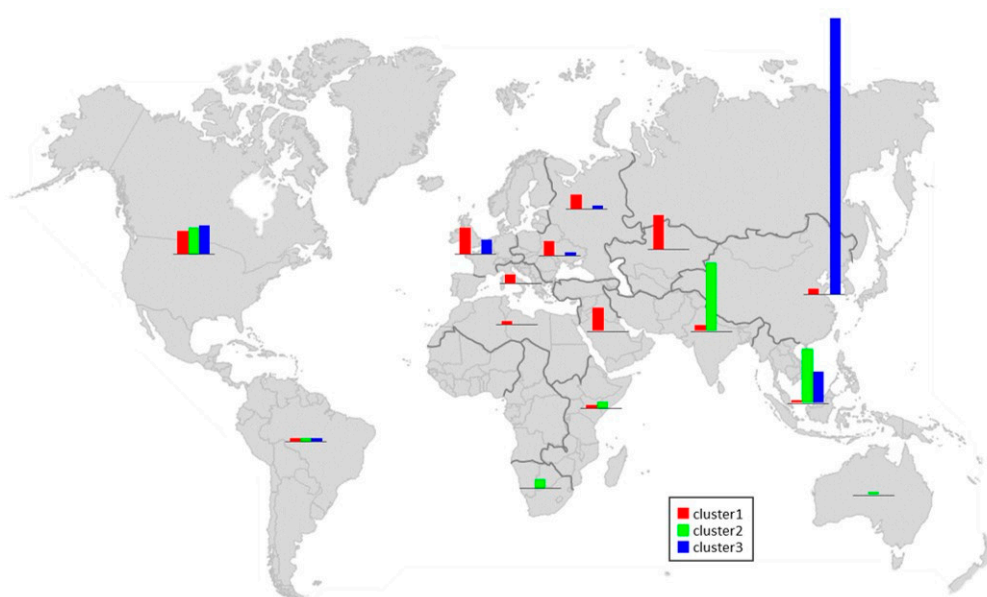
3. Results

3.1. Genotypes and Genetic Variation of the Cucumber Germplasm Collection

The GBS data were aligned to the genome sequence of *Cucumis sativus* L. var. *sativus* cv. 9930 [11]. The GBS genotyping of 264 accessions generated 947,205 SNPs after trimming a quality score of <30 (Table S2). The filtering of those SNPs with criteria of excluding the mono or the tri-allelic SNPs types, which included MAF > 0.05, SNP coverage > 0.6, and IF > 0.8, resulted in a set of 12,082 high-quality SNPs, which were evenly distributed on 7 chromosomes (Figure 1 and Table S2). Each SNP marker was named according to its physical position in the cucumber reference genome.

Figure 1. Genome-wide SNP distribution within a 1 Mb window size of a cucumber germplasm population.

3.2. Population Structure and Genetic Diversity of the Cucumber Germplasm Collection

To understand the structure of the cucumber germplasm collection, a structure analysis of ten independent runs was performed with K values from 1 to 10. Based on the genetic clustering algorithm model, the ΔK method determined that the best K is 3 for the SNP dataset. This result indicated that three subpopulations represent the best number of clusters for the cucumber germplasm collection used in this study. Cluster 1, cluster 2, and cluster 3 included 67, 61, and 136 accessions, respectively (Figure 2a and Table S3). It is interesting to note that at K = 4, which is not the best option of the clustering, cluster 3 in K = 3 was divided into breeding lines and germplasm, which is similar to the results seen in the PCA and the phylogenetic analysis (Figure 2).

Figure 2. Population structure and genetic relatedness of 264 cucumber accessions using genome-wide SNPs. (**a**) Population structure from the STRUCTURE analysis. ΔK reached its maximum value when K = 3 following the ad-hoc method. The subpopulations were grouped by Q value. (**b**–**d**) Genetic relatedness assessed by PCA. Each plot was distinguished in different colors based on the population structure (**b**), breeding line and germplasm (**c**), and variety group (**d**), respectively.

Evaluation of the geographical distribution showed that most of the accessions in cluster 1 were from Europe, Central Asia, and West Asia, whereas the accessions that belonged to cluster 2 were from South/East Asia and India, which is considered to be the center of the origin of cultivated cucumbers. Cluster 3 included accessions mostly from East Asia, which mostly included accessions from Korea, Japan, and China. Accessions from the North America were distributed in all three clusters (Figure 3). According to the Q value ratio, cluster 3 showed a homogeneous genetic background, whereas cluster 1 and 2 showed a slight admixture (Figure 2a and Table S3).

Figure 3. Distribution of 264 cucumber accessions in a population structure clusters according to their geographic location. The colors of the bar graph correspond to the clusters from the structure analysis. The length of each bar indicates the proportion of the included accessions.

The genetic difference among the three clusters was estimated by the population divergence (F_{ST}). The calculated F_{ST} value of 0.48 indicated the largest difference between the Eurasian (cluster 1) and the East Asian group (cluster 3). The second largest difference was identified between the Indian (cluster 2) and the East Asian group (cluster 3 and F_{ST} = 0.35). The smallest difference was found between the Eurasian (cluster 1) and the Indian group (cluster 2 and F_{ST} = 0.26). These results indicate that the genetic background of the East Asian group has a large difference compared to the other groups of cucumber germplasms.

3.3. Genetic Relatedness among the Population

The clustering pattern at optimal K from the structure analysis was consistent with the phylogenetic tree and the PCA results. The phylogenetic analysis using the weighted neighbor-joining clustering method separated the germplasm collection into four clades (Figure 4a). From the top, clades 1 to 3 were comprised mostly of the accessions in the cluster 3 in structure analysis, and clade 4 was comprised mostly of the accessions in cluster 1 and cluster 2 (Figure 4a).

Most of the breeding lines used in this study were clustered in clade 2. In this group, the diverse East-Asian breeding lines were strongly clustered. The Korean semi-white lines with a black spine (WB) and a white spine (WW) were closely related to the Korean solid green (HB) and the Chinese long green (CHS) (Figure 4b). Notably, some accessions from the WW were isolated in clade 3, which were placed in distinctive branches (Figure 4a). Meanwhile, the other breeding lines, such as the Beith Alpha (BAF and BAP), the American pickling (API and APD), the Parthenocarpic slicer (PSL), and the Taiwanese slicer (TAS), were divided into one separate group from the East-Asian breeding lines (Figure 4b and Table S4). Interestingly, some variety groups which included the CHS, the Thailand slicer (THS), and the Long Dutch green (EUR) showed stronger relations to the wild germplasms rather than the other breeding lines (Figure 4a). When performing the phylogenetic analysis with only the breeding lines, the above accessions were clustered in a different group from the others (Figure 4b).

A similar pattern of phylogenetic relationships was observed in the principal component analysis (PCA). When plotting the whole accessions in distance metrics, the first and the second axes explained 51.21% and 5.81% of the genotypic variance, respectively. Based on the PC1, the breeding lines were strongly grouped and separated from the wild germplasm accessions (Figure 2b). The wild germplasm

accessions were divided into three clusters according to the geographical distribution that was observed in the structure analysis (Figure 2a). Moreover, the breeding lines were closely grouped to each of the variety groups as shown in the phylogeny tree (Figure 2d and Table S4).
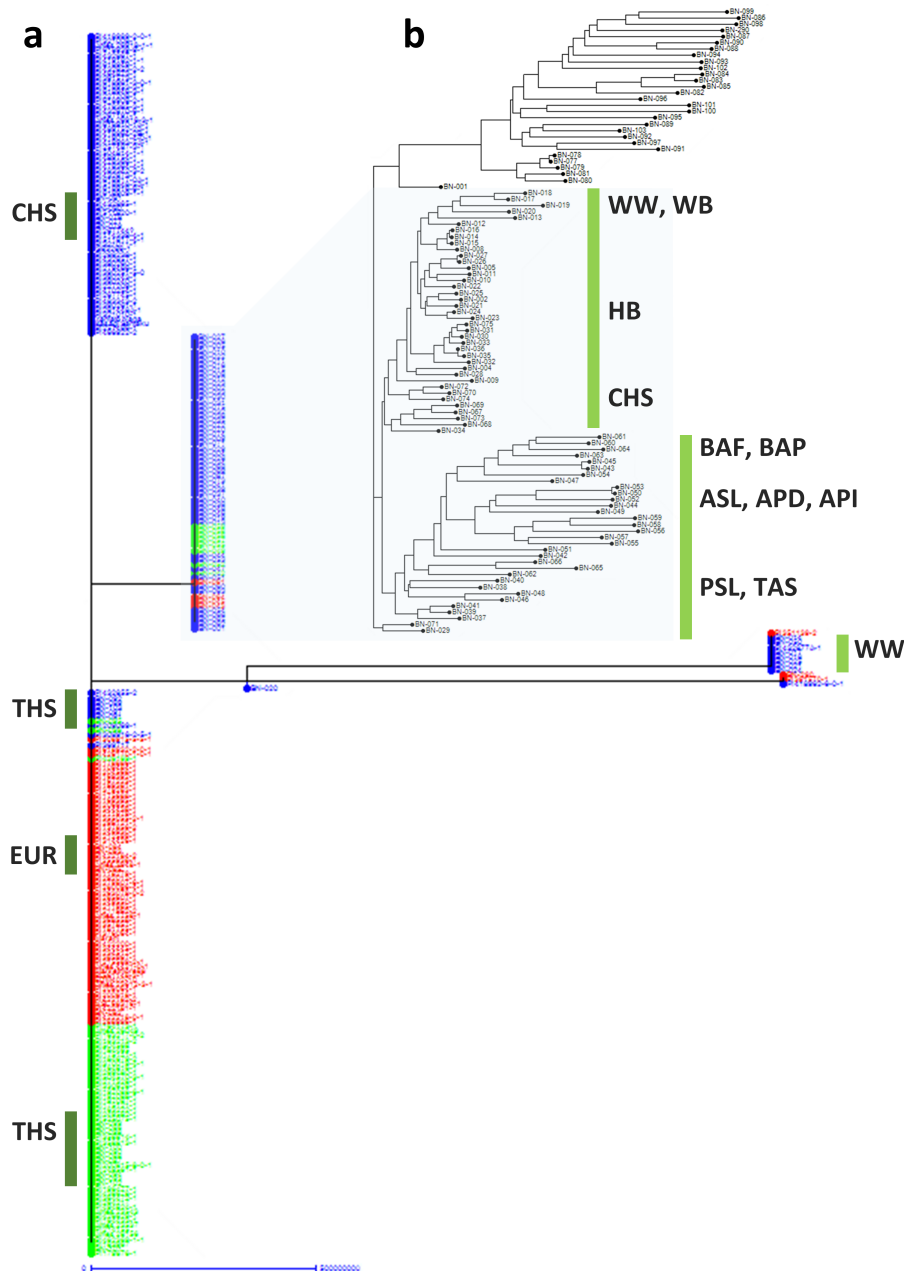
Figure 4. Genetic relatedness among 264 cucumber accessions assessed by the weighted neighboring joining method. (**a**) Phylogenic tree based on the whole germplasm collection; (**b**) phylogenic tree based on only 100 advanced breeding lines. Each branch was colored by the structure clusters. The colored boxes (green, and dark green) beside of the node denote the location of the breeding lines. The letters beside of the colored boxes indicate the abbreviation of the variety group (WB: Korean semi-white black spine, WW: Korean semi-white white spine, HB: Korean solid green, CHS: Chinese long green, THS: Thailand slicer, TAS: Taiwanese slicer, ASL: American slicer, PSL: Parthenocarpic slicer, APD: American pickling determinate, API: American pickling indeterminate, BAF: Beith Alpha open field, BAP: Beith Alpha parthenocarpic, EUR: Long Dutch green).

3.4. Evaluation of Powdery Mildew (PM) Resistance and Horticultural Traits

A total of 264 and 257 (except for the 7 weak accessions) accessions were evaluated for the resistance of powdery mildew (PM) in the 2018 and 2019 survey years, respectively. In 2018, 16 accessions showed high susceptibility. Five accessions—PI 438678, Ames 23537, PI 196477, Ames 23543, and PI 264217—showed strong resistance to the PM. However, most of the germplasms were susceptible to the PM and no accessions showed strong resistance in 2019. Even though there were different frequency distribution patterns in the disease resistance for each year, the normal distributions in the disease index were observed in both years (Figure 5a).
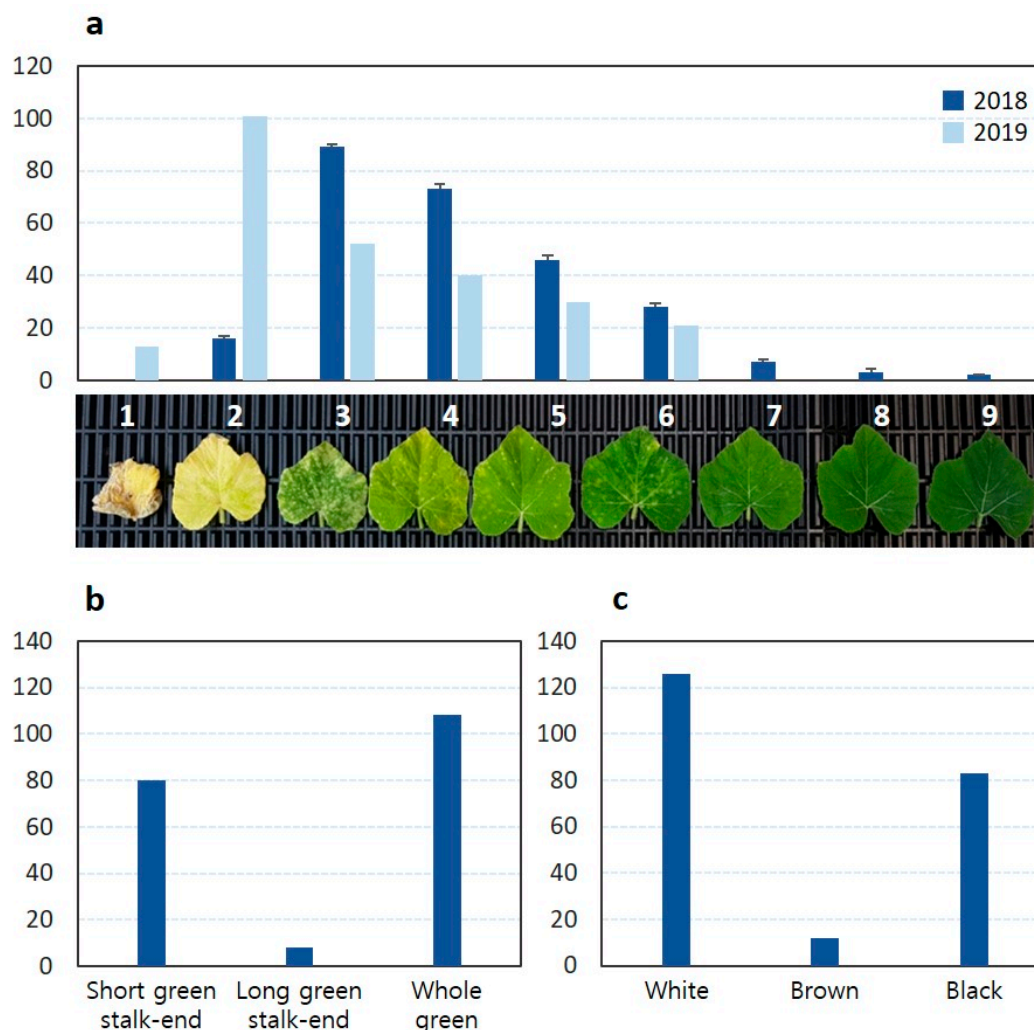
Figure 5. Frequency distribution of the phenotypes. (**a**) PM resistance; (**b**) fruit stalk end color; (**c**) spine color. The colors of bar graph correspond to the clusters from structure analysis. The blue and light blue bars correspond the phenotype data from 2018 and 2019, respectively. The upper bars denote the standard deviation. For the PM resistance, the phenotypic standard of each score were shown in images below the graph.

Two horticultural traits, the fruit's stalk-end color and the spine color, were evaluated in 2018 and 2019. The same results were identified from both evaluations, and among the 264 cucumber accessions, except for 68 accessions of the poor fruit, 40.8% of the entire germplasm showed whole green skin color, whereas 30.2% showed short green stalk-end with a light green skin color. Additionally, only 3% showed the long green stalk-end (Figure 5b and Figure S1). As for the spine color, except for

43 accessions where fruit was in bad condition, almost half of the accessions (47.5%) showed a white spine followed by a black spine (31.3%) and a brown spine (4.5%) (Figure 5c).

3.5. GWAS and Effective SNPs Influencing Horticultural Traits of Cucumber

The GWAS was performed for three horticultural traits—the PM resistance, the fruit's stalk-end color, and the spine color—and significant SNPs for each trait were identified. Five significant SNPs located at the 16 to 17 Mbp region on chromosome 5 were related to the PM resistance (Figure 6a). Among those SNPs, three SNPs, S5_16047445, S5_16080922, and S5_16623037, were commonly detected both in 2018 and 2019, and S5_16623037 showed the highest significance. In 2019, two additional significant SNPs, S5_17127151 and S5_17127216, were identified (Figure 6 and Table S5).
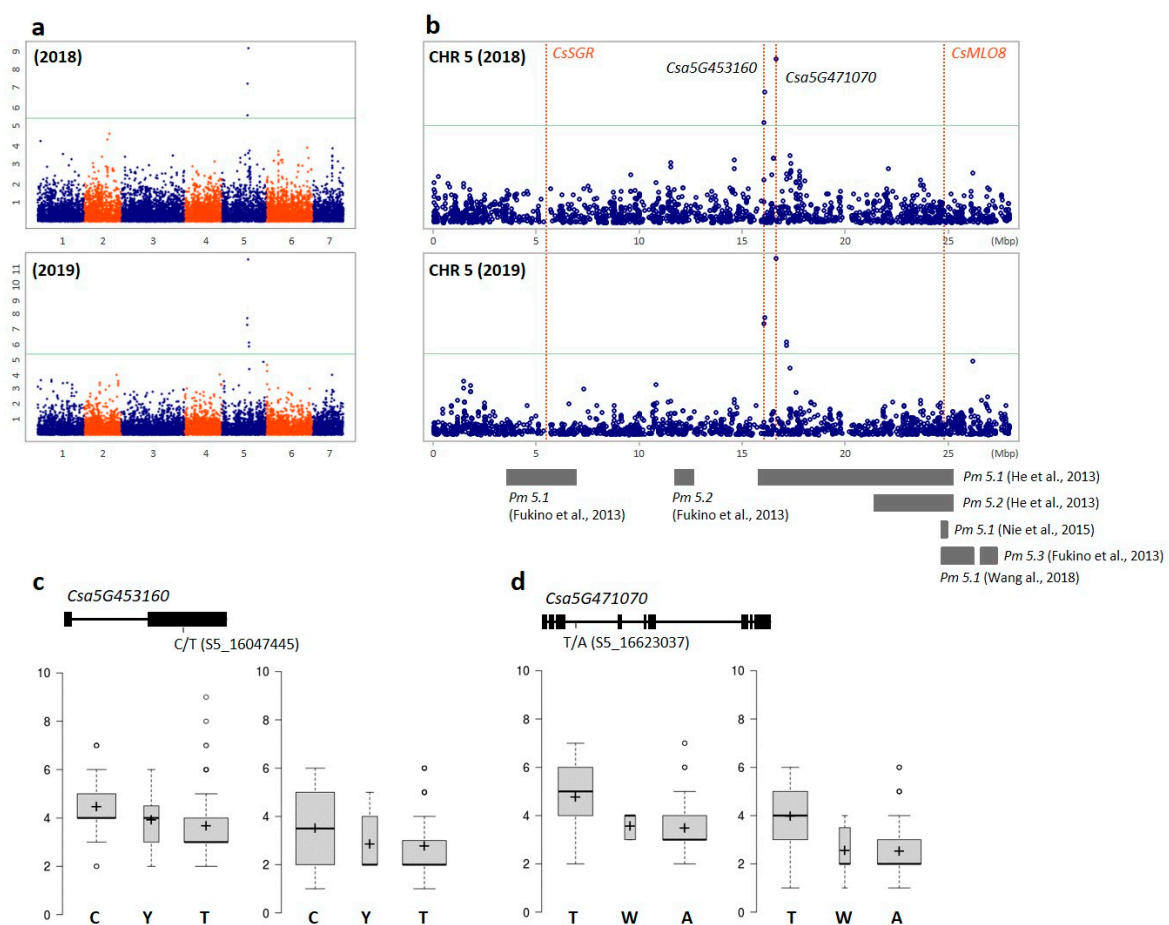
Figure 6. Significant SNPs and candidate genes correlating cucumber PM resistance from GWAS. The green horizontal line in the Manhattan plots indicate the significant threshold based on the Bonferroni correction. The position of the candidate genes and the reported genes are indicated by red dotted lines with the names. Under the chromosome-wide Manhattan plot, reported QTLs are represented by gray boxes. (**a**) Genome-wide association under 2018 and 2019 experimental year, respectively; (**b**) distribution of the significant SNPs and candidate genes on chromosome 5 with the reported genes and QTLs; (**c,d**) gene structure of candidate genes to PM with DNA polymorphism. Below the gene structure, boxplots show PM resistance index by years based on allelic differences of significant SNP; left from 2018, right from 2019; the width of each box is proportional to the square root of the number of accessions.

The fruit's stalk-end color showed an association with five SNPs. The most significant SNP was identified at 32.5 Mbp region on chromosome 3 ($-\log_{10}$ *p*-value = 7.59), which was followed by two

SNPs at 21.2 Mbp on chromosome 4, one SNP at 1.2Mbp on chromosome 1, and one SNP located at 14.1 Mbp on chromosome 3 (Figure S2 and Table S5).

The spine color was correlated with six SNPs. The most significant SNP was S5_7311032 ($-\log_{10}$ *p*-value = 7.87), which was followed by SNP S3_15236835, two SNPs within 17.7 Mbp, and the 21.1 Mbp region on chromosomes 2 and 6, respectively (Figure S2 and Table S5).

A total of 182, 558, and 509 genes were identified within the 1 Mbp regions around the significant SNPs for powdery mildew, the stalk end color, and the spine color, respectively. Among them, *Csa5G471070* and *Csa5G453160* were nominated as the candidate genes for the PM resistance as significant SNPs were located within the genes (Figure 6b). The candidate gene, which was the *Csa5G453160* linked to SNP s5_16047445, contained a significant SNP located inside the 2nd exon. The varieties carrying the C allele showed stronger resistance to the PM than those carrying the T allele in the two-year assessment (Figure 6c). The germplasm carrying the T allele of the SNP s5_16623037 located in the intron between the 6th and 7th exons of the *Csa5G471070* showed stronger resistance to the PM than those carrying the A allele for two years (Figure 6d).

Significant SNPs were found in two candidate genes, the *Csa1G006300* and the *Csa3G824850*, for fruit's stalk end color. The SNP S1_1181919 was located in the 5th intron of the *Csa1G006300*, and the S3_32581537 was located inside the 4th CDS region of the *Csa3G824850* (Figure S2).

For the spine color, all six detected SNPs were located in the genic regions. Two significant SNPs, the S2_17762942 and the S2_17762964, were located at the 1st exon of the *Csa2G368270*, and two SNPs, the S3_15236835 and the S5_7311032, were located in the intron of the *Csa5G453160* and the *Csa5G175680* on chromosome 2. Furthermore, two SNPs, the S6_21138663 and the S6_21138684, on chromosome 6 were linked to the *Csa6G448170* in the 2nd exon (Figure S2).

4. Discussion

In this study, we evaluated 264 *Cucumis* accessions that consisted of wild germplasms and various breeding lines of typical variety groups in order to understand the genetic diversity and the genetic relatedness especially between the breeding lines and the wild germplasms. A total of 100 advanced breeding lines of each typical variety group were assessed for genetic diversity and relatedness. For a comparison, 164 cucumber germplasm accessions were systematically sampled from the USDA NPGS based on the diversity. Among the germplasm accessions, 135 accessions are the same materials that were used for a previous study on genetic diversity, in which 76 out of 135 accessions were selected as a core collection [9]. This indicates that the germplasm accessions used in this study could be considered as a representative collection of the USDA cucumber collection.

Using these materials, the STRUCTURE analysis of the 264 *Cucumis* accessions revealed three distinct groups—group 1 (Indian and South Asia), group 2 (East Asia), and group 3 (others including Central/West Asia and Europe), according to the geographic origin and the domestication process, which are in accordance with the previous studies [9]. Based on the PCA and the phylogenetic analysis, we could confirm that the advanced breeding lines were distinguished from the other wild germplasm even though they consisted of various variety groups. The 15 variety groups used in this study have wide phenotypic variations in fruit morphology according to the typical consumption area. Therefore, we expected that there would be large genetic differences among the breeding lines. However, it is noteworthy that almost all the breeding lines were grouped together, which confirmed their small and narrow genetic background.

Within the breeding lines, two distinct groups were largely divided, and one of them was mainly cultivated in East Asia. One consisted of the WW, WB, HB, and CHS line, and the other was the slicer and pickling cucumbers that are grown in the USA and some European countries. This result was considered as a reflection of a narrow genetic background due to the highly integrated breeding within a limited area. Additionally, it supported the idea of the breeding process that cucumber domestication has been progressed in two directions—towards East Asia and towards another area [30].

The three traits assessed in this study—the PM, the spine color, and the fruit's stalk end color—are considered important characteristics in most of the local markets from the aspect of cultivation, storage, and consumption. In this study, we found genetic regions that are highly correlated with each trait by assessing the large scale of the germplasm collection as a natural population.

First of all, the PM resistance of 264 *Cucumis* accessions showed normal distribution, which was similar for two years, even though stronger resistance was observed in 2018. The difference in the disease index between the two years might be due to the different environmental conditions for each year. Based on the assessment of the disease index, five strong resistant germplasms to the PM were identified, and they were located in the lower part of the phylogeny tree (Figure 4a). The closest group of breeding line was the THS, even though the PM resistance of these variety groups was relatively lower than the wild accessions. Interestingly, among the WW-type breeding lines, the BN-019 showed a relatively strong PM resistance, and it seemed that it could be used as a potential resistant resource for the breeding of elite lines.

Using the two-year evaluation data, we obtained consistent results and detected three common significant SNPs in the 16 Mbp region on chromosome 5. This location is syntenic to the previously reported major QTLs, such as *pm 5.1* [21], *pm 5.1* [16], *pm 5.2* [44], *pm 5.2* [19], *pm 5.2* [20], and some genes related to the disease resistance (Figure 6b) [15]. Two significant SNPs for the PM were located within the genes, the *Csa5G471070* and the *Csa5G453160*, that encode the Anaphase-promoting complex subunit and short-chain dehydrogenase/reductase, respectively (Table 1).

Table 1. Summary of the candidate genes to powdery mildew resistance (PM), fruit stalk-end color (SDc), and spine color (SPc).

Traits	Gene ID	Protein Name	CHR	Start (bp)	End (bp)	Strand	Related SNPs
PM	*Csa5G453160*	Short-chain dehydrogenase/reductase	5	16,045,687	16,048,001	+	s5_16047445
	Csa5G471070	Anaphase-promoting complex subunit	5	16,621,524	16,631,782	−	s5_16623037
SDc	*Csa1G006300*	Response regulator 6	1	1,180,448	1,182,094	+	s1_1181919
	Csa3G824850	MYB transcription factor	3	32,580,286	32,581,685	+	s3_32581537
SPc	*Csa2G368270*	Pentatricopeptide repeat-containing protein	2	17,761,336	17,763,934	+	s2_17762942, s2_17762964
	Csa3G236570	Zinc-containing alcohol dehydrogenase quinone oxidoreductase	3	15,235,551	15,240,040	−	s3_15236835
	Csa5G175680	Pentatricopeptide repeat-containing protein	5	7,309,092	7,311,316	+	s5_7311032
	Csa6G448170	Thaumatin-like protein	6	21,137,853	21,139,047	+	s6_21138663, s6_21138684

Furthermore, one cyclin-like gene (*Csa5G488810*) and eight R genes (*Csa5G429990*, *Csa5G464830*, *Csa5G466350*, *Csa5G467390*, *Csa5G467900*, *Csa5G485190*, *Csa5G494390* and *Csa5G505160*) were identified around the 1 Mbp region of each significant SNP for the PM resistance. We should carefully look at the PM candidate genes from this study, since it was first reported in a cucumber's natural population that did not show a population-specific bias.

Two market traits, the SDc and the SPc, are particularly important in East Asian countries. According to the GWAS results, five and six significant SNPs were directly linked with two and four genes to the SDc and the SPc, respectively. For SDc, two candidate genes, *Csa1G006300* and *Csa3G824850*, which encode Response regulator 6 and MYB transcription factor were identified, respectively (Table 1). This is the first report to identify that the SNPs and the candidate genes are related to the cucumber fruit's stalk end color. These two genes are not directly related to the fruit color. However, in the Arabidopsis, the type-A RR genes play important roles in signaling transduction

in response to a wide range of biotic and abiotic stresses. The gene products function as cytokinin receptors and respond to stress negatively in the cytokinin signaling pathway [45]. The *Csa3G824850* gene shows homology to the *MYB16* in the *Arabidopsis*, and it has a 90% similarity to the MYB76-like isoform X2 in melons. The *MYB16* controls the epidermal cell morphogenesis in the petals or regulates the cuticle biosynthesis and the wax accumulation in the reproductive organs in the *Arabidopsis* [46,47].

Notably, all six significant SNPs for the spine color were located at genic regions encoding Pentatricopeptide repeat-containing protein, Zinc-containing alcohol dehydrogenase quinone oxidoreductase, Pentatricopeptide repeat-containing protein, and Thaumatin-like protein by *Csa2G368270*, *Csa3G236570*, *Csa5G175680*, and *Csa6G448170*, respectively (Table 1). According to the previous studies, the cucumber spine color is reported to be regulated by one or two strong dominant *B* genes. However, in this study, we could not identify a single strong gene, whereas significant SNPs on the chromosomes 2, 3, 5, and 6 were found. The most recently reported *HEUKCHEEM* gene was located at the 0.62 Mbp on chromosome 4 [30], whereas no relevant SNP was found on chromosome 4 in this study. In general, the PPR genes are involved in many cellular functions and biological processes, which include fertility restoration to the cytoplasmic male sterility (CMS) in plants. However, two PPR genes, the *Csa2G368270* and *Csa5G175680*, showed an association with spine color in this study, which showed homology to the *OTP90* and the *PCMP-H40*. These genes play a role in multiple site RNA editing in chloroplasts, such as C to U editing during the post-transcriptional step or editing of the site 7 of ndhB (ndhB-7) and site 5 of ndhD (ndhD-5) transcripts in the *Arabidopsis* [48,49]. According to the GO analysis, it was estimated that the *Csa3G236570* have molecular function on quinone reductase activity, and the *Csa6G448170* are related to the defense response.

In this study, we compared two groups of cucumber germplasms, wild germplasm and the advanced breeding lines, and found mutual genetic proximity. Even though the major consumption area or horticultural phenotypes of the breeding lines were significantly different, it was confirmed that the genetic backgrounds of the breeding lines were very narrow and clearly distinguished from the wild germplasm according to our results. Indeed, the breeding lines were divided into two groups that explained the direction of the cucumber domestication. By performing the GWAS, we identified 16 significant SNPs to three important traits (PM, SDc, SPc) and selected eight candidate genes. The identified SNPs and the candidate genes for three important traits (PM, SDc, SPc) via the GWAS performing are expected to be useful to develop molecular markers for cucumber breeding.

References

1. Che, G.; Zhang, X. Molecular basis of cucumber fruit domestication. *Curr. Opin. Plant Biol.* **2019**, *47*, 38–46. [CrossRef]

2. Shan, N.; Gan, Z.; Nie, J.; Liu, H.; Wang, Z.; Sui, X. Comprehensive Characterization of Fruit Volatiles and Nutritional Quality of Three Cucumber (*Cucumis Sativus* L.) Genotypes from Different Geographic Groups after Bagging Treatment. *Foods* **2020**, *9*, 294. [CrossRef]

3. Sebastian, P.; Schaefer, H.; Telford, I.R.H.; Renner, S.S. Cucumber (*Cucumis sativus*) and melon (*C. melo*) have numerous wild relatives in Asia and Australia, and the sister species of melon is from Australia. *Proc. Natl. Acad. Sci. USA* **2010**, *107*, 14269–14273. [CrossRef]

4. Yang, L.; Koo, D.H.; Li, Y.; Zhang, X.; Luan, F.; Havey, M.J.; Jiang, J.; Weng, Y. Chromosome rearrangements during domestication of cucumber as revealed by high-density genetic mapping and draft genome assembly. *Plant J.* **2012**, *71*, 895–906. [CrossRef]

5. Bo, K.; Ma, Z.; Chen, J.; Weng, Y. Molecular mapping reveals structural rearrangements and quantitative trait loci underlying traits with local adaptation in semi-wild Xishuangbanna Cucumber (*Cucumis sativus* L. *var*. *xishuangbannanesis* Qi et Yuan). *Theor. Appl. Genet.* **2015**, *128*, 25–39. [CrossRef]

6. Gaut, B.S.; Seymour, D.K.; Liu, Q.; Zhou, Y. Demography and its effects on genomic variation in crop domestication. *Nat. Plants* **2018**, *4*, 512–520. [CrossRef] [PubMed]

7. Osipowski, P.; Pawełkowicz, M.; Wojcieszek, M.; Skarzyńska, A.; Przybecki, Z.; Pląder, W. A high-quality cucumber genome assembly enhances computational comparative genomics. *Mol. Genet. Genomics* **2020**, *295*, 177–193. [CrossRef] [PubMed]

8. Qi, J.; Liu, X.; Shen, D.; Miao, H.; Xie, B.; Li, X.; Zeng, P.; Wang, S.; Shang, Y.; Gu, X.; et al. A genomic variation map provides insights into the genetic basis of cucumber domestication and diversity. *Nat. Genet.* **2013**, *45*, 1510–1515. [CrossRef] [PubMed]

9. Wang, X.; Bao, K.; Reddy, U.K.; Bai, Y.; Hammar, S.A.; Jiao, C.; Wehner, T.C.; Ramírez-Madera, A.O.; Weng, Y.; Grumet, R.; et al. The USDA cucumber (*Cucumis sativus* L.) collection: Genetic diversity, population structure, genome-wide association studies, and core collection development. *Hortic. Res.* **2018**, *5*. [CrossRef] [PubMed]

10. Renner, S.S. A valid name for the Xishuangbanna gourd, a cucumber with carotene-rich fruits. *PhytoKeys* **2017**, *94*, 87–94. [CrossRef]

11. Huang, S.; Li, R.; Zhang, Z.; Li, L.; Gu, X.; Fan, W.; Lucas, W.J.; Wang, X.; Xie, B.; Ni, P.; et al. The genome of the cucumber, *Cucumis sativus* L. *Nat. Genet.* **2009**, *41*, 1275–1281. [CrossRef]

12. Li, Q.; Li, H.; Huang, W.; Xu, Y.; Zhou, Q.; Wang, S.; Ruan, J.; Huang, S.; Zhang, Z. A chromosome-scale genome assembly of cucumber (*Cucumis sativus* L.). *Gigascience* **2019**, *8*, 1–10. [CrossRef] [PubMed]

13. Cavagnaro, P.F.; Senalik, D.A.; Yang, L.; Simon, P.W.; Harkins, T.T.; Kodira, C.D.; Huang, S.; Weng, Y. Genome-wide characterization of simple sequence repeats in cucumber (*Cucumis sativus* L.). *BMC Genom.* **2010**, *11*, 569. [CrossRef]

14. Rubinstein, M.; Katzenellenbogen, M.; Eshed, R.; Rozen, A.; Katzir, N.; Colle, M.; Yang, L.; Grumet, R.; Weng, Y.; Sherman, A.; et al. Ultrahigh-Density Linkage Map for Cultivated Cucumber (*Cucumis sativus* L.) Using a Single-Nucleotide Polymorphism Genotyping Array. *PLoS ONE* **2015**, *10*, e0124101. [CrossRef]

15. Zhang, P.; Zhu, Y.; Wang, L.; Chen, L.; Zhou, S. Mining candidate genes associated with powdery mildew resistance in cucumber via super-BSA by specific length amplified fragment (SLAF) sequencing. *BMC Genom.* **2015**, *16*, 1058. [CrossRef]

16. Wang, Y.; VandenLangenberg, K.; Wen, C.; Wehner, T.C.; Weng, Y. QTL mapping of downy and powdery mildew resistances in PI 197088 cucumber with genotyping-by-sequencing in RIL population. *Theor. Appl. Genet.* **2018**, *131*, 597–611. [CrossRef]

17. Chen, Q.; Yu, G.; Wang, X.; Meng, X.; Lv, C. Genetics and Resistance Mechanism of the Cucumber (*Cucumis sativus* L.) Against Powdery Mildew. *J. Plant Growth Regul.* **2020**. [CrossRef]

18. Sakata, Y.; Kubo, N.; Morishita, M.; Kitadani, E.; Sugiyama, M.; Hirai, M. QTL analysis of powdery mildew resistance in cucumber (*Cucumis sativus* L.). *Theor. Appl. Genet.* **2006**, *112*, 243–250. [CrossRef]

19. Fukino, N.; Yoshioka, Y.; Sugiyama, M.; Sakata, Y.; Matsumoto, S. Identification and validation of powdery mildew (*Podosphaera xanthii*)-resistant loci in recombinant inbred lines of cucumber (*Cucumis sativus* L.). *Mol. Breed.* **2013**, *32*, 267–277. [CrossRef]

20. He, X.; Li, Y.; Pandey, S.; Yandell, B.S.; Pathak, M.; Weng, Y. QTL mapping of powdery mildew resistance in WI 2757 cucumber (*Cucumis sativus* L.). *Theor. Appl. Genet.* **2013**, *126*, 2149–2161. [CrossRef] [PubMed]

21. Nie, J.; He, H.; Peng, J.; Yang, X.; Bie, B.; Zhao, J.; Wang, Y.; Si, L.; Pan, J.S.; Cai, R. Identification and fine mapping of *pm 5.1*: A recessive gene for powdery mildew resistance in cucumber (*Cucumis sativus* L.). *Mol. Breed.* **2015**, *35*, 1–13. [CrossRef]

22. Yoshioka, Y.; Sakata, Y.; Sugiyama, M.; Fukino, N. Identification of quantitative trait loci for downy mildew resistance in cucumber (*Cucumis sativus* L.). *Euphytica* **2014**, *198*, 265–276. [CrossRef]

23. Liu, P.N.; Miao, H.; Lu, H.W.; Cui, J.Y.; Tian, G.L.; Wehner, T.C.; Gu, X.F.; Zhang, S.P. Molecular mapping and candidate gene analysis for resistance to powdery mildew in *Cucumis sativus* Stem. *Genet. Mol. Res.* **2017**, *16*, 1–10. [CrossRef]

24. Xu, X.; Yu, T.; Xu, R.; Shi, Y.; Lin, X.; Xu, Q.; Qi, X.; Weng, Y.; Chen, X. Fine mapping of a dominantly inherited powdery mildew resistance major-effect QTL, *pm 1.1*, in cucumber identifies a 41.1 kb region containing two tandemly arrayed cysteine-rich receptor-like protein kinase genes. *Theor. Appl. Genet.* **2016**, *129*, 507–516. [CrossRef] [PubMed]

25. Wang, Y.; Tan, J.; Wu, Z.; VandenLangenberg, K.; Wehner, T.C.; Wen, C.; Zheng, X.; Owens, K.; Thornton, A.; Bang, H.H.; et al. STAYGREEN, STAY HEALTHY: A loss-of-susceptibility mutation in the *STAYGREEN* gene provides durable, broad-spectrum disease resistances for over 50 years of US cucumber production. *New Phytol.* **2019**, *221*, 415–430. [CrossRef]

26. Sheng, Y.; Pan, Y.; Li, Y.; Yang, L.; Weng, Y. Quantitative trait loci for fruit size and flowering time-related traits under domestication and diversifying selection in cucumber (*Cucumis sativus*). *Plant Breed.* **2020**, *139*, 176–191. [CrossRef]

27. Guo, Y.; Gao, M.; Liang, X.; Xu, M.; Liu, X.; Zhang, Y.; Liu, X.; Liu, J.; Gao, Y.; Qu, S.; et al. Quantitative Trait Loci for Seed Size Variation in Cucurbits—A Review. *Front. Plant Sci.* **2020**, *11*, 1–12. [CrossRef]

28. Wang, Y.; Bo, K.; Gu, X.; Pan, J.; Li, Y.; Chen, J.; Wen, C.; Ren, Z.; Ren, H.; Chen, X.; et al. Molecularly tagged genes and quantitative trait loci in cucumber with recommendations for QTL nomenclature. *Hortic. Res.* **2020**, *7*. [CrossRef]

29. Li, Y.; Wen, C.; Weng, Y. Fine mapping of the pleiotropic locus *B* for black spine and orange mature fruit color in cucumber identifies a 50 Kb region containing a R2R3-MYB transcription factor. *Theor. Appl. Genet.* **2013**, *126*, 2187–2196. [CrossRef]

30. Zhang, C.; Win, K.T.; Kim, Y.C.; Lee, S. Two types of mutations in the *HEUKCHEEM* gene functioning in cucumber spine color development can be used as signatures for cucumber domestication. *Planta* **2019**, *250*, 1491–1504. [CrossRef]

31. Pierce, L.K.; Wehner, T.C. Review of Genes and Linkage Groups in Cucumber. *HortScience* **1990**, *25*, 605–615. [CrossRef]

32. Walters, S.A.; Shetty, N.V.; Wehner, T.C. Segregation and Linkage of Several Genes in Cucumber. *J. Am. Soc. Hortic. Sci.* **2001**, *126*, 442–450. [CrossRef]

33. Han, K.; Lee, H.Y.; Ro, N.Y.; Hur, O.S.; Lee, J.H.; Kwon, J.K.; Kang, B.C. QTL mapping and GWAS reveal candidate genes controlling capsaicinoid content in *Capsicum*. *Plant Biotechnol. J.* **2018**, *16*, 1546–1558. [CrossRef]

34. Lee, H.-Y.; Ro, N.-Y.; Patil, A.; Lee, J.-H.; Kwon, J.-K.; Kang, B.-C. Uncovering Candidate Genes Controlling Major Fruit-Related Traits in Pepper via Genotype-by-Sequencing Based QTL Mapping and Genome-Wide Association Study. *Front. Plant Sci.* **2020**, *11*, 1–15. [CrossRef] [PubMed]

35. Li, H.; Durbin, R. Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics* **2010**, *26*, 589–595. [CrossRef]

36. Depristo, M.A.; Banks, E.; Poplin, R.; Garimella, K.V.; Maguire, J.R.; Hartl, C.; Philippakis, A.A.; Del Angel, G.; Rivas, M.A.; Hanna, M.; et al. A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat. Genet.* **2011**, *43*, 491–501. [CrossRef]

37. Hubisz, M.J.; Falush, D.; Stephens, M.; Pritchard, J.K. Inferring weak population structure with the assistance of sample group information. *Mol. Ecol. Resour.* **2009**, *9*, 1322–1332. [CrossRef]

38. Pritchard, J.K.; Stephens, M.; Donnelly, P. Inference of Population Structure Using Multilocus Genotype Data. *Genetics* **2000**, *155*, 945–959. [CrossRef] [PubMed]

39. Evanno, G.; Regnaut, S.; Goudet, J. Detecting the number of clusters of individuals using the software Structure: A simulation study. *Mol. Ecol.* **2005**, *14*, 2611–2620. [CrossRef]

40. Perrier, X.; Jacquemoud-Collet, J.P. DARwin Software. Available online: http://darwin.cirad.fr/darwin (accessed on 6 November 2020).

41. Stacklies, W.; Redestig, H.; Scholz, M.; Walther, D.; Selbig, J. pcaMethods—A bioconductor package providing PCA methods for incomplete data. *Bioinformatics* **2007**, *23*, 1164–1167. [CrossRef]

42. Lipka, A.E.; Tian, F.; Wang, Q.; Peiffer, J.; Li, M.; Bradbury, P.J.; Gore, M.A.; Buckler, E.S.; Zhang, Z. GAPIT: Genome association and prediction integrated tool. *Bioinformatics* **2012**, *28*, 2397–2399. [CrossRef]

43. Bonferroni, C.E. Teoria statistica delle classi e calcolo delle probabilita. *Pubbl. R Ist. Sup. Sci. Econ. Commer. Fir.* **1936**, *8*, 3–62.

44. Zhang, S.P.; Liu, M.M.; Miao, H.; Zhang, S.Q.; Yang, Y.H.; Xie, B.Y.; Gu, X.F. QTL mapping of resistance genes to powdery mildew in cucumber (*Cucumis sativus* L.). *Sci. Agric. Sin.* **2011**, *44*, 3584–3593.

45. Tran, L.S.P.; Shinozaki, K.; Yamaguchi-Shinozaki, K. Role of cytokinin responsive two-component system in ABA and osmotic stress signalings. *Plant Signal. Behav.* **2010**, *5*, 148–150. [CrossRef]

46. Baumann, K.; Perez-Rodriguez, M.; Bradley, D.; Venail, J.; Bailey, P.; Jin, H.; Koes, R.; Roberts, K.; Martin, C. Control of cell and petal morphogenesis by R2R3 MYB transcription factors. *Development* **2007**, *134*, 1691–1701. [CrossRef] [PubMed]

47. Oshima, Y.; Shikata, M.; Koyama, T.; Ohtsubo, N.; Mitsuda, N.; Ohme-Takagi, M. MIXTA-Like Transcription Factors and WAX INDUCER1/SHINE1 Coordinately Regulate Cuticle Development in *Arabidopsis* and *Torenia fournieri*. *Plant Cell* **2013**, *25*, 1609–1624. [CrossRef] [PubMed]

48. Malbert, B.; Burger, M.; Lopez-Obando, M.; Baudry, K.; Launay-Avon, A.; Härtel, B.; Verbitskiy, D.; Jörg, A.; Berthomé, R.; Lurin, C.; et al. The Analysis of the Editing Defects in the *dyw2* Mutant Provides New Clues for the Prediction of RNA Targets of Arabidopsis E+-Class PPR Proteins. *Plants* **2020**, *9*, 280. [CrossRef]

49. Okuda, K.; Chateigner-Boutin, A.L.; Nakamura, T.; Delannoy, E.; Sugita, M.; Myouga, F.; Motohashi, R.; Shinozaki, K.; Small, I.; Shikanai, T. Pentatricopeptide Repeat Proteins with the DYW Motif Have Distinct Molecular Functions in RNA Editing and RNA Cleavage in *Arabidopsis* Chloroplasts. *Plant Cell* **2009**, *21*, 146–156. [CrossRef] [PubMed]

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.