Article

Development of SNP Marker Sets for Marker-Assisted Background Selection in Cultivated Cucumber Varieties

Eun Su Lee [†] , Hee-Bum Yang [†], Jinhee Kim, Hye-Eun Lee, Ye-Rin Lee and Do-Sun Kim ^{*}

Vegetable Research Division, National Institute of Horticultural and Herbal Science, Rural Development Administration, Wanju 55365, Korea; lus4434@korea.kr (E.S.L.); yhb0130@korea.kr (H.-B.Y.); sayzinni@korea.kr (J.K.); helee72@korea.kr (H.-E.L.); lyr1219@korea.kr (Y.-R.L.)

^{*} Correspondence: greenever@korea.kr

[†] These authors contributed equally to this work.

Abstract: Marker-assisted background selection is a powerful molecular tool that can enhance breeding efficiency through the analysis of a large number of markers representing the entire genomic background for precise selection. In the present study, the transcriptomes of 38 cucumber inbred lines with diverse traits were sequenced for single nucleotide polymorphism (SNP) mining for practical application to commercial cucumber breeding. A total of 62,378 high-quality SNPs were identified, of which 2462 SNPs were chosen based on the stringent filtering parameters. Finally, 363 evenly distributed common background selection markers (BMs) were developed and validated through polymorphism analysis and phylogenetic analysis using breeding materials with different genetic backgrounds; 327 out of 363 common BMs were useful for background selection. Moreover, the results of the phylogenetic analysis carried out using 50 selected core BMs were consistent with those for 327 common BMs. However, when the genotypes of breeding materials belonging to only the Baekdadagi-type were analyzed, the 327 common BMs showed a significant reduction in polymorphisms within the biased genomic locations. To address this issue, 59 highly polymorphic markers were selected as Baekdadagi BMs, as they showed better selection ability for the Baekdadagi-type. The 327 common BMs developed in the present study will enable efficient marker-assisted background selection in cucumber. Additionally, to reduce the genotyping cost, we suggested an alternative background selection strategy using both evenly distributed core BMs and biased Baekdadagi BMs for the improvement of commercial cucumber breeding programs.

Keywords: cucumber; genetic background selection; SNP mining; molecular markers; NGS analysis

Citation: Lee, E.S.; Yang, H.-B.; Kim, J.; Lee, H.-E.; Lee, Y.-R.; Kim, D.-S. Development of SNP Marker Sets for Marker-Assisted Background Selection in Cultivated Cucumber Varieties. *Agronomy* **2022**, *12*, 487. <https://doi.org/10.3390/agronomy12020487>

Academic Editor: Pasquale Tripodi

Received: 17 January 2022

Accepted: 11 February 2022

Published: 16 February 2022

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

Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Cucumber, *Cucumis sativus* L. ($2n = 2x = 14$), an important cucurbitaceous crop, is one of the ten major vegetables in the world. Approximately 70% of the world's cucumber production occurs in Asia, including China, Turkey, Iran, and Russia [1]. Cucumber is a semitropical vegetable crop native to southern Asia. Four major types of cucumber, the Huánan-type (southern China), Huábei-type (northern China), European-type, and a Huánan × Huábei-type, are consumed worldwide [2]. Among these types, the Baekdadagi-type and Chwicheong-type belong to the Huábei group, and the Gasi-type belongs to the Huánan group [2]. The Baekdadagi-type cucumber is cultivated mostly in northeastern Asian countries, including Korea, China, and Japan. It is also a popular cucumber variety on the Korean market and is consumed as a preparation in various side dishes [2]. The unique characteristic of the Baekdadagi-type cucumber is its whitish, light-green skin color. In contrast, most of the other cucumber types consumed in China and Europe have a green or dark-green skin color [1]. The upper part of the stalk region and the lower part of the blossom region of the fruit of the Baekdadagi-type cucumbers have a green skin color, which gradually turns into white-green as it moves down to the center of the fruit.

Baekdadagi-type cucumber lines mostly have long fruits with slightly bumpy skin. More than 80% of cucumber cultivation comes from Baekdadagi-type cucumber varieties in Korea, and these varieties represent the largest percentage of an annual production in the Korean cucumber market [3].

Cucumber is a diploid vegetable crop with a genome size of 367 Mbp [4]. High-quality reference genomes and high-throughput genotyping methods have greatly reduced the time needed for the identification of useful genes in crop plants [5,6]. Advancements in next-generation sequencing (NGS) and genotyping technologies have enabled the rapid and cost-effective development of DNA markers, including single nucleotide polymorphisms (SNPs). SNPs are biallelic and provide abundant genetic variation due to their high frequency and even distribution throughout the genome [7]. Therefore, it is now possible to scan an entire genome of any organism at high marker densities to identify associations among individual markers [8]. Several studies utilizing SNP marker sets in cucumber have recently been published [2,9–12]. These studies focused on germplasm accession speciation, variety identification, genome-wide association studies (GWASs), and geographic evolution. These studies have proven the efficiency of the SNP-based background selection markers (BMs) in cucumber and the usefulness of genotyping studies using the Fluidigm platform [2,9,10].

Marker-assisted selection (MAS), using trait-linked markers, is widely employed breeding technology in many crops [13]. Furthermore, background selection breeding techniques like marker-assisted backcrossing (MABC), which uses many genomic markers in combination with single or few markers for target traits such as secondary metabolites, disease resistance, and tolerance against environmental stress, have been reported in several crop breeding programs. A capsinoid-associated gene was introgressed in pepper breeding lines by MABC with 412 SNPs used for MAS [14]. Improvement of Indian groundnut regarding foliar disease resistance and high levels of oleic acid was accomplished by MABC using 58 K SNPs used for MAS based on QTLs for both traits [15]. Drought-tolerant bread wheat has also been developed by MABC with 600 simple sequence repeats (SSRs) and MAS based on targeted QTL-associated markers [16].

In this study, we employed inbred lines belonging to six cucumber types with different morphological characteristics that are used as commercial cucumber breeding materials in Korea. We identified useful SNPs and developed 363 evenly distributed common BMs for practical use in cucumber commercial breeding programs. Moreover, 50 representative core BMs and 59 biased Baekdadagi BMs were also validated, which significantly reduced the genotyping cost. Hence, we suggested an alternative type-specific background selection strategy by combining the two types of marker sets.

2. Materials and Methods

2.1. Plant Materials

A set of 38 inbred lines, provided by Won-nong Seed Co., Ltd. (company A; Ansung, Korea), was selected based on important traits such as spine color, heat tolerance, female flower rates, powdery mildew resistance, flowering node number, and skin color. These lines were divided into six types: Tropical, Pickle, Mini, Beit Alpha, Baekdadagi, and Chwicheong (Table S1). These 38 inbred lines (hereafter referred to as “inbred group A”) were used for the transcriptome sequencing. Twenty-eight lines among those in inbred groups A; the parents of 31 F₁ combinations (hereafter referred to as “parental group”) classified into seven types: Tropical, Pickle, Mini, Beit Alpha, Baekdadagi, Chwicheong, and Gasi, obtained from company A; and 29 inbred lines (hereafter referred to as “inbred group B”) provided by NongwooBio Co., Ltd. (company B; Suwon, Korea) were used for the validation of the markers developed in this study. These three kinds of breeding materials were examined as test materials. Finally, Baekdadagi-type lines from the parental group and inbred group B were selected to develop a Baekdadagi-type-specific marker set (hereafter referred to as “Baekdadagi group”).

2.2. RNA Extraction and Transcriptome Sequencing

For transcriptome sequencing, RNA from each sample was isolated from young leaf tissues 20 days after germination. The GeneAll Hybrid-RTM kit (GeneAll Biotechnology Co., Ltd., Daejeon, Korea) was used for total RNA extraction following the manufacturer's protocol. The purity and concentration of the RNA were determined by a Nano Drop analyzer (NanoDrop Technologies, Wilmington, DE, USA). An Agilent 2100 Bioanalyzer (Agilent Technologies, Palo Alto, CA, USA) was used for quality control of the RNA. We selected high-quality RNA samples showing RNA integrity number (RIN) values over 7 and 28S/18S rRNA ratios over 1. cDNA synthesis and RNA-seq library construction were performed at the Macrogen Inc. based on the protocols for the Illumina HiSeq2500 (Illumina, San Diego, CA, USA), and 101 bp paired-end sequencing was performed.

2.3. Transcriptome Analysis for the Identification and Filtering of Useful SNPs for Breeding

The sequencing data from each lane were merged for further analysis. The raw fastq-formatted short-reads data generated from the Illumina HiSeq2500 were trimmed to remove low-quality reads using Trimmomatic 0.3238 [17] according to Kim et al. [18]. Read sequences below a Phred score of 20 were eliminated. Subsequently, the trimmed reads were mapped to the cucumber reference genome sequence retrieved from the database of "Chinese long cucumber ver. 2" at <http://cucurbitgenomics.org> (accessed on 12 December 2021) [4,5] using TopHat 2.0.1339 [19] and Burrows–Wheeler Aligner (BWA 0.7.8-r455) [20] according to Kim et al. [18].

Further read group addition and sorting were performed by the Picard tool 1.112 package [21], and genotyping and unification of the SNPs were accomplished by employing GATK3.1 [22–25]. For SNPs filtration, SAMtools 0.1.840 [26] along with in-house customized scripts were used. This pipeline focused on selecting SNPs by comparing all the sequences from the 38 inbred group A individuals one by one to avoid reference bias. The identified crude SNPs were filtered based on the following criteria: (i) short read depth > 10, (ii) homozygous and diallelic SNPs, (iii) minor allele frequency (MAF): polymorphism information content (PIC) > 0.36, (iv) 1:1 segregation ratio, and (v) distance between flanking SNPs. The PIC values of the markers were calculated by the formula $PIC = 1 - \sum_{i=1}^n P_i^2 - \sum_{i=1}^{n-1} \sum_{j=i+1}^n 2P_i^2 P_j^2$ (P_i is the allele frequency of the i allele, and P_j is the allele frequency of the j allele) using an in-house Python script [27].

2.4. Development and Validation of SNP Markers

Selected SNPs were used to develop Fluidigm SNP type™ assays (Fluidigm Corp., South San Francisco, CA, USA) to identify genetic background selection markers (common BMs and Baekdadagi BMs in this study). Specific target amplification primers (STA), locus-specific primers (LSP), and allele-specific primers (ASP1 & 2) were designed for each SNP (Table S2). Genotype analysis of Fluidigm markers was performed according to the manufacturer's protocol using the Fluidigm EP1™ System. The SNP genotyping analysis was performed by Fluidigm SNP Genotyping Analysis software (Fluidigm Corp., South San Francisco, CA, USA; version 4.5.1). KASP™ assays (LGC Biosearch™ Technologies, Teddington, UK) were designed to develop core genetic background selection markers (core BMs in this study). Two allele-specific forward primers and a reverse primer for each SNP were designed by LGC's own program. Genotype analysis of KASP assays was performed by the Foundation of Agri. Tech. Commercialization & Transfer (Iksan, Korea). To validate the SNP markers, total genomic DNA was isolated from young green leaves of cucumbers using the CTAB method according to Hwang et al. [28]. The purity and quantity of the DNA were estimated by a NanoDrop (NanoDrop Technologies, Wilmington, DE, USA).

2.5. Phylogenetic Analysis and Drawing the Marker Density Diagram

Cluster analysis of marker genotyping results was analyzed using the neighbor-joining algorithm. A phylogenetic tree of cucumbers was constructed using the DARWIN 6.0 soft-

ware (<http://darwin.cirad.fr/darwin> (accessed on 17 January 2022)) [29]. Tree construction was based on the dissimilarity matrix calculated with the Manhattan index. The analysis was performed with 1000 bootstrap replicates from the generated genetic distance matrix. The marker density diagram of the selected SNPs was drawn by an in-house Python script based on scalable vector graphics (SVG) language.

2.6. Functional Analysis of Putative Gene Bearing SNPs for the Background Markers

Functional analysis was carried out to identify genes containing the SNPs found in common BMs. Putative genes overlapping the physical positions of SNPs for the common BM were selected from the “Chinese long cucumber ver. 2” database at <http://cucurbitgenomics.org> (accessed on 26 December 2021) [4,5]. A BLAST search using the blastn function at <https://blast.ncbi.nlm.nih.gov> (accessed on 26 December 2021) was performed to find putative genes containing the SNPs when the putative genes were not found in the database [30]. Gene Ontology (GO) terms for the selected putative genes were predicted by Blast2GO software (BioBam Bioinformatics S.L., Valencia, Spain) [31]. First, local BLAST was performed to find high scoring pairs (HSPs) for each putative gene against the nucleotide collection database (accessed on 26 December 2021) downloaded at <https://blast.ncbi.nlm.nih.gov> with parameters, “-word_size 6” and “-threshold 21” [32]. Then all the data were merged into Blast2GO software. GO terms were categorized and visualized by using WEGO 2.0 software at <https://wego.genomics.cn> (BGI Genomics, Shenzhen, China; accessed on 12 January 2022) [33]. Coding sequence region (CDS) data from “Chinese long cucumber genome ver. 3” [6] was downloaded, and GO analysis was conducted to compare the distribution pattern of GO terms originating from markers in this study with whole putative genes in cucumber. The GO pattern comparison was accomplished using WEGO 2.0 software.

3. Results

3.1. Functional Analysis of Putative Genes Bearing SNPs for the Background Markers

Transcriptome sequencing of 38 inbred group A individuals was performed for the identification of SNPs. Transcriptome sequencing resulted in an average of approximately 3,602,733,209 bp of data and 35,990,890 reads. The number of cucumber genes was expected to be 23,248, with an estimated mean gene length of 3213 bp [5]. The total physical length of the coding region was calculated to be approximately 75 Mbp after multiplication of the number of genes by the mean gene length. By this calculation, the average transcriptome coverage was approximately 48-fold. The GC content of the sequences ranged from 43.81 to 49.93%. The Phred quality scores indicating the quality of the sequencing results had values ranging from 98.5 to 99.29% for Q20, whereas, for Q30, the values were 94.85–97.52%. The trimmed reads for each inbred line were aligned to the *Cucumis sativus* reference genome [5]. The number of filtered SNPs and percentage for every filtering step are listed in Table 1. Initially, 62,378 reliable SNPs were discovered after the first filtering step (sequence depth ≥ 10). Next, we obtained 51,435 homozygous diallelic SNPs with a PIC value of 0.36 or higher, which implies that at least 15 out of 38 inbred group A individuals are expected to be distinguishable (Table 1). The segregation ratio and distance among flanking SNPs were also considered for the following SNP selection procedure. A strict criterion of a segregation ratio close to 1:1, for example 19:19 or 18:20, was applied to identify highly polymorphic SNPs; then, SNPs within the 60 bp window were also eliminated to prevent the overlap of markers for neighboring SNPs. After filtering based on these criteria, a total of 2462 SNPs were retained (Figure 1A and Table 1).

Table 1. The number of SNPs retained after each filtering procedure.

Filtering Criteria	No. of Remaining SNPs	No. of Filtered SNPs	Filtering Percentage (%) ¹
SNPs (reads depth ≥ 10)	62,378	-	-
Homozygous/Diallelic	58,436	3942	6.3
MAF (PIC > 0.35)	51,435	7001	12
Segregation ratio (1:1)	4985	46,450	90.3
Flanking SNP (>60 bp)	2462	2523	50.6
No. of BM ² markers	371	2091	84.9

¹ Filtering percentage (%): (no. of filtered SNP/no. of remaining SNPs in former criteria) $\times$ 100. ² BMs: background selection markers.

Figure 1. Marker density of SNPs and marker sets: (A) 2462 SNPs filtered by five criteria, (B) 363 common BMs, (C) 50 core BMs, and (D) 59 Baekdadi BMs. Bars in squares represent the physical locations of SNPs and markers on chromosomes.

3.2. Development and Validation of Common BMs in the Fluidigm Platform

A total of 371 evenly distributed SNPs throughout all seven cucumber chromosomes were selected for Fluidigm marker development. The range of SNP densities in each chromosome was recorded as being between 1.6 and 2.8 SNPs/Mb with an average of 2.3 SNPs/Mb (Table 2). The number of SNPs per chromosome differed among the seven chromosomes. Chromosome 3 contained the highest number of SNPs, while chromosome 1 displayed the lowest number of SNPs. Among 371 SNPs, 8 were eliminated because of their inability to be converted into Fluidigm markers according to the manufacturer's program. Finally, 363 common BMs for genetic background selection using the Fluidigm platform were developed (Table 2 and Figure 1). A polymorphism survey of those markers was performed using 28 inbred group A individuals used for transcriptome sequencing. A total of 327 markers showed polymorphism, while 36 markers were excluded due to the lack of a genotype call (Figure S1).

Table 2. The number of SNPs retained after each filtering procedure.

Chromosome	Size (Mbp)	No. of SNPs/Markers (Density ¹)				
		Selected SNPs	Common BMs ^{2,3}	Polymorphic Common BMs	Core BMs ⁴	“Baekdadagi” BMs ⁵
Chr1	22.7	37 (1.6)	37 (1.6)	36 (1.6)	6 (0.3)	0 (0.0)
Chr2	20.6	48 (2.3)	47 (2.3)	41 (2.0)	7 (0.3)	11 (0.5)
Chr3	39.0	78 (2.0)	76 (1.9)	71 (1.8)	9 (0.2)	18 (0.5)
Chr4	22.6	60 (2.7)	60 (2.7)	56 (2.5)	8 (0.4)	8 (0.4)
Chr5	10	22 (2.2)	22 (2.2)	19 (1.9)	5 (0.5)	1 (0.1)
Chr6	27.4	74 (2.7)	71 (2.6)	59 (2.2)	8 (0.3)	5 (0.2)
Chr7	18.8	52 (2.8)	50 (2.7)	45 (2.4)	7 (0.4)	16 (0.9)
Total	161	371 (2.3)	363 (2.3)	327 (2.0)	50 (0.3)	59 (0.4)

¹ Density: the number of SNPs or markers/chromosome size (Mbp). ² BMs: background selection markers.

³ Common BMs: SNP marker set for background selection in a variety of cucumber types. ⁴ Core BMs: a portion of the SNP marker set selected from common BMs. ⁵ Baekdadagi BMs: SNP marker set for background selection in Baekdadagi-type cucumber.

The phylogenetic analysis was performed using the genotyping data for 327 SNPs in 28 inbred group A individuals, which divided the lines into three clades and clustered them into six groups according to their types (Figure 2). Lines belonging to the Chwicheong-type and Baekdadagi-type were clustered into a single clade, and the Mini-type lines were separated into two clades. Pickle-type lines, except C05, Beit-Alpha-type lines, and Tropical-type lines were aggregated into a single clade. The Pickle-type and Beit-Alpha-type subgroups were closer together, and the Chwicheong- and Baekdadagi-types were also grouped near each other (Figure 2).

Figure 2. The phylogenetic tree of 28 inbred group A individuals constructed based on common BMs.

On the other hand, a polymorphism survey using 29 inbred group B individuals was performed to validate the application of common BMs to different breeding material groups with diverse genetic backgrounds. Approximately 99.1% of the 327 markers, except for 3, showed polymorphisms in inbred group B (breeding materials of company B). However,

the PIC value for the common BMs in inbred group B was significantly reduced compared to that in inbred group A (Table 3). The PIC values of 29.6% of the polymorphic markers were more than 0.3, which indicated that these markers were able to distinguish more than a quarter of inbred group B individuals.

Table 3. Comparison of the PIC values for polymorphic markers in breeding materials.

Statistical Analysis	363 Common BMs ¹			50 Core BMs		59 Baekdadagi BMs
	28 Inbred Group A Individuals	29 Inbred Group B Individuals	36 Baekdadagi Group Individuals	62 Parental Group Individuals	36 Baekdadagi Group Individuals	36 Baekdadagi Group Individuals
Min of PICs	0.27	0.03	0.05	0.07	0.05	0.26
Max of PICs	0.375	0.375	0.37	0.37	0.37	0.375
Average of PICs	0.37	0.22	0.21	0.25	0.10	0.35
Standard deviation of PICs	0.08	0.12	0.12	0.08	0.11	0.03
25% percentile of PICs	0.37	0.12	0.10	0.20	0.00	0.33
Median of PICs	0.37	0.21	0.18	0.23	0.00	0.36
75% percentile of PICs	0.37	0.34	0.33	0.31	0.18	0.37
No. of polymorphic markers	327	324	189	48	22	59
Ratio of markers (0.1 > PIC > 0) ²	0%	18.2%	33.3%	4.2%	31.8%	-
Ratio of markers (0.2 > PIC > 0.1)	0%	26.9%	17.5%	22.9%	18.2%	-
Ratio of markers (0.3 > PIC > 0.2)	0.9%	25.3%	18.0%	43.8%	22.7%	8.5%
Ratio of markers (PIC > 0.3)	99.1%	29.6%	31.2%	29.2%	27.3%	91.5%

¹ BMs: background selection markers. ² Ratio of markers = (number of markers)/(total number of polymorphic markers) × 100.

In summary, a total of 327 markers showed polymorphism in at least one line of the breeding materials from the two companies A and B (inbred groups A and B). Conversely, 36 markers were not usable for genotyping in either of the breeding materials (Figure S1). A total of 324 markers showed polymorphism in both breeding materials, and the PIC values for more than one-quarter of the markers were above 0.3 in breeding materials with different genetic backgrounds. Therefore, these markers can be successfully applied to any of the cucumber breeding materials.

3.3. Development and Validation of KASP Markers as Core BMs

Out of the 324 SNPs distributed evenly on each chromosome, a total of 50 SNPs were identified and converted to KASP markers (5–9 SNPs per chromosome) (Table 2 and Figure 1C). Phylogenetic analysis of 28 inbred group A individuals using 50 core BMs showed results consistent with those of 363 common BMs, except for minor differences (Figure S2). From these results, we concluded that core BMs can be used for background selection as a small set for cucumber breeding, which will ultimately reduce the genotyping cost.

For the validation test, the core BMs were used to evaluate 62 parental group individuals from company A, which were classified into six types. This panel also included another cucumber type, Gasi, which was not subjected to transcriptome analysis. Two of the fifty markers did not show polymorphism; however, the PIC value of the other 48 markers in the 62 parental lines was similar to those of 324 markers in inbred lines from the company B (Table 3).

Furthermore, the heterozygosity of the 62 parental lines was calculated. Heterozygous genotypes were scored in more than a quarter of the 62 parental lines for 13 markers, CsaSPT011, CsaSPT012, CsaSPT013, CsaSPT017, CsaSPT020, CsaSPT021, CsaSPT022, CsaSPT023, CsaSPT024, CsaSPT031, CsaSPT046, CsaSPT049, and CsaSPT050 (Tables S2 and S3). Four markers among them, CsaSPT017, CsaSPT021, CsaSPT022, and CsaSPT043, showed heterozygous genotypes in more than half of the 31 F₁ hybrids derived from 62 parental lines. These highly heterozygous markers may be employed not only for genetic background selection but also for seed purity test.

3.4. The Development of Type-Specific BMs for Baekdadagi Breeding Materials

Cucumber breeding goals in Korea mainly focus on the Baekdadagi-type. Therefore, it is important to develop a highly polymorphic marker set for breeding materials belonging to this type. Genotypes of Baekdadagi group individuals, including 20 inbred group A and 16 inbred group B individuals, were analyzed using 327 common BMs. Only 189 markers out of 327 showed polymorphism. The PIC values were also reduced compared to those of other breeding materials (Table 3). From these results, it can be assumed that this obvious reduction in marker polymorphism and PIC values may be due to the narrowing of genetic variation through the artificial selection of Baekdadagi-type breeding materials. Of the 363 common BMs, a total of 59 markers having PIC values higher than 0.26 were selected as Baekdadagi BMs, with an average PIC value of 0.35 in the Baekdadagi group (Figure 1D, Tables 3, S2 and S4). Baekdadagi BMs, representing genomic regions with higher polymorphism in the Baekdadagi group, were not distributed evenly across the genome; instead, they were aggregated into limited regions (Table 2). Chromosomes 3 and 7 contained the highest number of polymorphic markers, while chromosomes 1 and 5 did not harbor any polymorphic markers or had only one marker (Table 2).

When we tested the core BMs on Baekdadagi group via a phylogenetic analysis, the grouping patterns of the core BMs were not consistent with those of the common BMs; therefore, the core BMs did not represent the complete genomic background of the Baekdadagi group (Figure S3B). This result might be due to the loss of polymorphisms in approximately 40% of the core BMs and the low PIC values of the remaining polymorphic markers compared to other breeding lines (Table S3). On the other hand, the grouping pattern of newly developed Baekdadagi BMs was more consistent with that of the common BMs than that of the core BMs (Figure S3C). Taking all these results together, it could be concluded that the genetically diverse genomic regions of the Baekdadagi group are biased; therefore, every marker in the common BM set will not be required in Baekdadagi-type breeding programs. Preferably, Baekdadagi BMs would provide practically applicable information for the genetic background selection of Baekdadagi-type lines. Finally, when we merged genotype data from the core BMs and Baekdadagi BMs and performed phylogenetic analysis, the pattern was highly consistent with that of the common BMs (Figure S3D). It is supposed that the whole genome coverage of the core BMs and the higher polymorphism resolution of the Baekdadagi BMs worked complementarily to improve the background selection power.

3.5. Computational Annotation of Putative Genes Bearing SNPs in Common BM

All background selection markers in the present study were developed based on transcriptome data, thus these markers might be related to coding regions in the cucumber genome. Comparative analysis revealed that a total of 318 putative genes matched with 363 common BMs in the “Chinese long cucumber ver. 2” genome database [4] regarding the physical location of putative genes and SNPs employed for common BM development (Table S2). A BLAST search was carried out for the 45 unannotated common BMs, and 41 putative genes were found, while no putative genes were found for 4 common BMs. Finally, excluding 39 duplicated genes, 320 putative genes were selected for further GO analysis. Of the 320 putative genes, 245 had matches in the GO database. Most GO terms were classified as intracellular organelles, including membrane-bound organelles and non-membrane-bound organelles and protein-containing complexes for the cellular component ontology (Figure 3A). Additionally, a large majority of GO terms were assigned to the hydrolase activity, catalytic activity on protein, transferase activity, and nucleic-acid-binding categories for the molecular function ontology. Finally, GO terms were also involved in nitrogen compounds, organic substances, and primary and cellular metabolic processes in the biological process ontology (Figure 3A).

Figure 3. The comparative GO analysis results of putative genes containing SNPs for common BMs (blue) and whole cucumber putative genes in cucumber, as reported by Li et al. (2019) [6] (orange). (A) GO assignment of putative genes derived from common BMs at GO level 4. (B) Comparative analysis of 363 common BMs and whole putative genes in the cucumber genome based on GO assignment pattern at GO level 4. All GO terms presented in (B) satisfied the Pearson chi-square test at p value < 0.05.

Whole putative genes predicted in the “Chinese long cucumber database ver. 3” database [6] were employed to compare the GO patterns with those of common BMs. The ratios of putative genes found for common BMs were significantly higher than those of whole cucumber putative genes in the following categories: cytoplasm, intracellular organelle, and protein-containing complex categories in the cellular component ontology; heterocyclic compound binding, including nucleic acid binding, and catalytic activity, including hydrolase activity, in the molecular function ontology; and cellular metabolic process in the biological process ontology. The ratios of all putative genes were higher than those of common BMs in some GO categories; however, the differences were not significant under the Pearson chi-square test.

4. Discussion

In this study, we selected 38 genetically diverse inbred group A individuals classified into six different types and identified a large number of SNPs in more than 60k of transcriptome analysis data. A total of 2462 reliable SNPs were selected by filtering based on strict criteria, and 363 SNPs were selected to develop common BMs for genetic background selection. A total of 324 markers showed polymorphisms in both breeding materials from two seed companies. Phylogenetic analysis was performed, and most of the inbred lines were aggregated into the expected groups. Furthermore, 50 evenly distributed markers that represent the overall genomic background were selected and employed as core BMs, and 59 markers that were highly polymorphic in Baekdadagi breeding materials, called the Baekdadagi group, were selected as Baekdadagi BMs. The common BMs can also be used for the genetic background selection, such as marker-assisted backcrossing (MABC) of Baekdadagi-type lines and other cucumber types. A marker set developed by a robust background genotyping technique such as Fluidigm could shorten the breeding period significantly via the early selection procedure, which could possibly lead to precise breeding. Eventually, we suggested an alternative practical background selection strategy using core BMs combined with Baekdadagi BMs, type-specific BMs, at relatively low cost.

High-throughput SNP marker sets have been used to scan the whole genome background of several horticultural crops, and numerous SNP marker sets for cucumber have also been developed [2,9,10,12]. A new SNP genotyping technology called Target SNP-seq was developed to efficiently analyze cucumber genotypes [12]. This tool combines the advantages of multiplex PCR amplification and high-throughput sequencing. The DNA fingerprint of 261 cucumber varieties was analyzed by Target SNP-seq using 163 SNPs. The core set of 24 SNPs was also developed to distinguish 261 cucumber varieties. Other studies analyzed the two newly developed SNP sets in various cucumber accessions and F₁ hybrids [2,10]. The 151 Fluidigm SNP assays were able to distinguish 280 cucumber accessions collected from the four continents [2]. The 96-SNP core set was also developed to facilitate F₁ cultivar identification. A total of 88 commercial F₁ hybrids from various crosses were successfully distinguished by this core set [10]. The genetic diversity of the worldwide cucumber germplasm collection (264 accessions) was also analyzed using genome-wide SNP markers (>12,082 SNPs) [9]. These markers were obtained by a genotyping-by-sequencing (GBS) approach. These SNP marker sets are mainly focused on distributing diverse cucumber types, which can be simply classified by their phenotypes.

The germplasm of a species contains a set of genetic material that has maximum broad genetic variations; therefore, the SNP mining using accessions has the advantage of obtaining the most abundant SNPs [34,35]. However, many genetic and phenotypic variations in wild varieties have been lost during the domestication process, which reduced the frequency of unfavorable alleles and increased the frequency of alleles that benefitted humankind [36]. Whole genome sequencing technologies enable us to study diversity at the DNA level. In peach, the first reduction in nucleotide diversity occurred during the original domestication (4000–5000 years ago), after which a second reduction occurred recently (16th–19th century) [35]. The modern breeding process also contributed to the reduction in genetic diversity. A total of 75 Canadian hard red spring wheat varieties released from 1845 to 2004 were used for the genetic diversity study, which revealed that some new genetic varieties were introduced while many more genetic varieties were lost [37]. This compelling evidence suggests that SNPs found in germplasm populations can be lost in modern breeding materials. Our research strategy of SNP mining using inbred lines aimed to minimize this risk. The total number of SNPs must be less than the number of SNPs found in the germplasm; however, the rates of retaining variation in breeding materials should be higher. Our results showed that 324 (99.1%) out of 327 polymorphic Fluidigm markers, common BMs, from group A showed polymorphisms in group B, supporting this hypothesis. These markers are expected to maintain polymorphisms in breeding materials with shared genetic backgrounds of groups A and B.

Interestingly, only 60% of the polymorphic common BMs in 28 inbred group A individuals maintained their polymorphisms in Baekdadagi-type inbred lines, indicating that approximately 40% of the genomic regions in the cucumber population belonging to Baekdadagi-type lost genetic diversity (Table 3). After domestication, the cucumber population was differentiated into three groups [9,11], and several types were separated by modern breeding [1]. During these genetic events, it is supposed that the genetic erosion of diversity might have occurred by unexplored causes, such as adaptation to local climate and artificial selection based on food preferences and local cultural traditions [2]. This result is consistent with molecular variance (AMOVA) analysis results that described the genetic divisions that occurred when cucumber types were differentiated, with the genetic variance present among the types being higher than that within each type [2].

Our results showed that another feature of Baekdadagi-specific genetically diverse genomic regions is maintenance of genetic diversity; however, this diversity was not distributed evenly across the cucumber genomes. These results are consistent with selective sweep events which happened during domestication and modern breeding in mung bean and lettuce [38,39]. Putative selective sweep regions of both crops were also not distributed evenly. Evenly distributed SNPs have been selected in most of the studies for genetic background selection, such as GWAS and species/varieties identification in cucumber [2,9,10,12]. Those studies assumed that evenly distributed SNPs can represent the whole genome, and this assumption is based on the idea that genetically important genes are distributed evenly across the genome. However, by considering only Baekdadagi-type breeding materials, genetic background selection using biased type-specific markers, for example, Baekdadagi BMs, will be very important because genetically important regions are biased in the Baekdadagi-type population, which is supported by the biased marker density and the phylogenetic analysis of Baekdadagi-type inbred lines based on three BM sets in our study (Figure S3A–C).

Furthermore, the size of regions with marker coverage was also a reason implying the need for type-specific markers. Most of the evenly distributed markers lost their genotype-discriminating ability in two or three generations because each marker covered a wide genetic region [14]. In contrast, type-specific markers cover relatively narrow type-specific genetically diverse regions; hence, there is a possibility of maintaining genotype-discriminating ability at a much higher rate than what is achieved with evenly distributed markers. Practically, genetic background selection using combinations of evenly distributed marker sets and type-specific biased markers will be a better alternative strategy. For example, the Baekdadagi-type breeding team can select both core BMs and Baekdadagi BMs for their breeding program (Figure S3D). The whole genomic background can be analyzed by core BMs and Baekdadagi-type-specific genetic regions can be analyzed by Baekdadagi BMs. The total number of markers to be genotyped under this strategy would be less than the number in the common BM set; however, these markers might be sufficient to distinguish and select individuals in the same generation with reduced cost.

Computational annotation of the common BMs was performed, and the results were compared with those of all putative genes in the cucumber genome (Table S2 and Figure 3). A large number of putative genes involved in the common BMs were assigned to several different GO categories, such as intracellular organelles, including membrane-bound organelles and non-membrane-bound organelles for the cellular component ontology; catalytic activity, including hydrolase activity, and heterocyclic compound complexes, including nucleic acid binding, for the molecular function ontology; and cellular processes, including cellular metabolic processes, for the biological process ontology. Furthermore, the ratio of assigned putative genes for these GO categories was significantly higher than that of all putative genes. We detected polymorphic SNPs located in putative genes; however, further study will be required to determine the possible involvement of these markers and genes in biological functions and practical breeding.

Even though highly strict filtering based on five criteria was applied to select SNPs for the development of 363 Fluidigm markers, common BMs, 36 markers were excluded

due to a high ratio of heterozygous genotypes, bias toward one of the genotypes, a high ratio of no calls, and wide dispersion of the dots in scatter plot (Figure S1). These markers have corresponding genes because they were developed by employing SNPs found in the transcriptome, and studies of those genes could provide a clue to understand why 36 of the markers were not able to correctly detect SNPs. The failure of Fluidigm marker genotype calling was reported previously, and repetitive sequences and SNPs with low coverage were assumed to be the main causes of failure [40]. Annotation data for the genes bearing SNPs within the 36 excluded markers were retrieved, and a BLAST search was performed [4,5]. It was determined that 10 putative genes, Csa1G573590, Csa3G038110, Csa3G535630, Csa5G153030, Csa6G008780, Csa6G324840, Csa6G366530, Csa6G483300, Csa7G071350, and Csa7G071630, have at least one additional homolog in cucumber (Table S5). Two putative genes, Csa6G052090 and Csa6G078520, were single-copy genes in cucumber, but homologs were found in the corresponding genes of other species. The polymerase chain reaction (PCR) step was required for the analysis of the Fluidigm markers, and in this process, not only specific DNA targets but also non-targeted DNA can be amplified together if the target gene has a homolog in the genome. Diagnostic research using NGS technologies for human diseases suggests that a minimum coverage of 25-fold is required for the precise detection of heterozygous genotype [41]. However, the minimum depth of the SNP filtering criteria in our research was 10-fold. If the number of SNPs is sufficient for selection, increasing the minimum depth standard by more than 10 would be helpful for increasing the ratio of markers able to correctly call an SNP in the Fluidigm platform.

Our results showed that the Baekdadagi group has a distinct pattern of genetic diversity. Other types of cucumber may have their own patterns, so diversity analysis of genome for each type is required to develop type-specific markers. Universal background marker sets for every cucumber type, such as common BMs, are the best first step toward realizing background selection breeding; however, many more type-specific markers will be required for precise breeding. With reduced NGS analysis costs, abundant SNPs have been identified in many species. However, the relatively higher cost of genotyping technology delays the application of these abundant SNPs to practical plant breeding; therefore, the ability to select SNPs containing abundant genetic diversity within target breeding materials is essential. We hope that our strategies, including core inbred line selection, strict SNP filtering, and the selection of effective core SNPs and type-specific SNPs, will guide the identification of commercially available SNPs for effective background selection markers in cucumber breeding programs at low cost.

5. Conclusions

The background selection technique using a large set of molecular markers covering the whole genome is a very useful tool for the quick development inbred lines in commercial seed companies. Various high-throughput genotyping platforms have been developed; however, relatively higher genotyping costs are a barrier to the extensive application of this technique to plant breeding programs. In this study, we isolated a large number of SNPs by analyzing the transcriptomes of cucumber inbred lines. A total of 363 common background selection markers (BM) evenly distributed on the cucumber genome were developed based on SNPs selected under strict criteria. We showed that genotyping using Baekdadagi-specific BMs and core BMs, which are a small set of common BMs, was adequate for practical use in Baekdadagi-type cucumber breeding programs. This result proved that reductions in the genotyping cost can be achieved by developing type-specific BMs.

Supplementary Materials: The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/agronomy12020487/s1>, Figure S1: Scatter plots of excluded markers; Figure S2: Phylogenetic tree of 28 inbred lines constructed by 50 core BMs genotypes; Figure S3: Phylogenetic analysis of Baekdadagi-type inbred lines; Table S1: The characteristics of 38 cucumber inbred lines; Table S2: Summary of background markers used in this study; Table S3: PIC values of core BMs; Table S4: PIC values of Baekdadagi BMs; Table S5: Copy number variation of genes from BMs that failed to call a genotype.

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

Funding: This study was supported by the National Institute of Horticultural & Herbal Science (NIHHS), Rural Development Administration (RDA), Korea (Project No. PJ01185401).

Data Availability Statement: Sequencing reads have been deposited in the National Center for Biotechnology Information (NCBI) Sequencing Read Archive (PRJNA780020). The other data and materials presented in this study are mentioned in the main text as well as in the Supplementary Files; further data will be provided on request from the corresponding author.

Acknowledgments: H.-B.Y. was supported by the RDA Research Associate Fellowship Program of the National Institute of Horticultural and Herbal Science, Rural Development Administration, Korea. We are very grateful to Won-nong Seed Co., Ltd. And NongwooBio Co., Ltd. For providing seeds of the inbred lines and F₁ hybrids. Special thanks to our colleague Irfan for his work on English editing.

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

References

1. Naegel, R.P.; Wehner, T.C. Genetic Resources of Cucumber. In *Genetics and Genomics of Cucurbitaceae*; Grumet, R., Katzir, N., Garcia-Mas, J., Eds.; Plant Genetics and Genomics: Crops and Models; Springer International Publishing: Cham, Switzerland, 2016; Volume 20, pp. 61–86; ISBN 978-3-319-49330-5.
2. Park, G.; Choi, Y.; Jung, J.-K.; Shim, E.-J.; Kang, M.; Sim, S.-C.; Chung, S.-M.; Lee, G.P.; Park, Y. Genetic Diversity Assessment and Cultivar Identification of Cucumber (*Cucumis Sativus* L.) Using the Fluidigm Single Nucleotide Polymorphism Assay. *Plants* **2021**, *10*, 395. [CrossRef] [PubMed]
3. Center for Agricultural Outlook. Available online: <https://aglook.krei.re.kr> (accessed on 10 September 2021).
4. Li, Z.; Zhang, Z.; Yan, P.; Huang, S.; Fei, Z.; Lin, K. RNA-Seq Improves Annotation of Protein-Coding Genes in the Cucumber Genome. *BMC Genom.* **2011**, *12*, 540. [CrossRef] [PubMed]
5. 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]
6. 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*, giz072. [CrossRef] [PubMed]
7. Jenkins, S.; Gibson, N. High-Throughput SNP Genotyping. *Comp. Funct. Genom.* **2002**, *3*, 57–66. [CrossRef] [PubMed]
8. Ganai, M.W.; Altmann, T.; Röder, M.S. SNP Identification in Crop Plants. *Curr. Opin. Plant Biol.* **2009**, *12*, 211–217. [CrossRef] [PubMed]
9. Lee, H.Y.; Kim, J.G.; Kang, B.C.; Song, K. 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. *Agronomy* **2020**, *10*, 1736. [CrossRef]
10. Park, G.; Sim, S.-C.; Jung, J.-K.; Shim, E.-J.; Chung, S.-M.; Lee, G.P.; Park, Y. Development of Genome-Wide Single Nucleotide Polymorphism Markers for Variety Identification of F₁ Hybrids in Cucumber (*Cucumis Sativus* L.). *Sci. Hortic.* **2021**, *285*, 110173. [CrossRef]
11. Wang, X.; Bao, K.; Reddy, U.K.; Bai, Y.; Hammar, S.A.; Jiao, C.; Wehner, T.C.; Ramirez-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*, 64. [CrossRef]
12. Zhang, J.; Yang, J.; Zhang, L.; Luo, J.; Zhao, H.; Zhang, J.; Wen, C. A New SNP Genotyping Technology Target SNP-Seq and Its Application in Genetic Analysis of Cucumber Varieties. *Sci. Rep.* **2020**, *10*, 5623. [CrossRef]
13. Hasan, N.; Choudhary, S.; Naaz, N.; Sharma, N.; Laskar, R.A. Recent Advancements in Molecular Marker-Assisted Selection and Applications in Plant Breeding Programmes. *J. Genet. Eng. Biotechnol.* **2021**, *19*, 128. [CrossRef] [PubMed]
14. Jeong, H.S.; Jang, S.; Han, K.; Kwon, J.K.; Kang, B.C. Marker-Assisted Backcross Breeding for Development of Pepper Varieties (*Capsicum Annuum*) Containing Capsinoids. *Mol. Breed.* **2015**, *35*, 226. [CrossRef]
15. Shasidhar, Y.; Variath, M.T.; Vishwakarma, M.K.; Manohar, S.S.; Gangurde, S.S.; Sriswathi, M.; Sudini, H.K.; Dobariya, K.L.; Bera, S.K.; Radhakrishnan, T.; et al. Improvement of Three Popular Indian Groundnut Varieties for Foliar Disease Resistance and High Oleic Acid Using SSR Markers and SNP Array in Marker-Assisted Backcrossing. *Crop J.* **2020**, *8*, 1–15. [CrossRef]
16. Rai, N.; Bellundagi, A.; Kumar, P.K.C.; Kalasapura Thimmappa, R.; Rani, S.; Sinha, N.; Krishna, H.; Jain, N.; Singh, G.P.; Singh, P.K.; et al. Marker-Assisted Backcross Breeding for Improvement of Drought Tolerance in Bread Wheat (*Triticum Aestivum* L. Em Thell). *Plant Breed.* **2018**, *137*, 514–526. [CrossRef]
17. Bolger, A.M.; Lohse, M.; Usadel, B. Trimmomatic: A Flexible Trimmer for Illumina Sequence Data. *Bioinformatics* **2014**, *30*, 2114–2120. [CrossRef] [PubMed]

18. Kim, J.; Manivannan, A.; Kim, D.-S.; Lee, E.-S.; Lee, H.-E. Transcriptome Sequencing Assisted Discovery and Computational Analysis of Novel SNPs Associated with Flowering in *Raphanus Sativus* In-Bred Lines for Marker-Assisted Backcross Breeding. *Hortic. Res.* **2019**, *6*, 120. [CrossRef] [PubMed]
19. Kim, D.; Pertea, G.; Trapnell, C.; Pimentel, H.; Kelley, R.; Salzberg, S.L. TopHat2: Accurate Alignment of Transcriptomes in the Presence of Insertions, Deletions and Gene Fusions. *Genome Biol.* **2013**, *14*, R36. [CrossRef]
20. Li, H.; Durbin, R. Fast and Accurate Short Read Alignment with Burrows–Wheeler Transform. *Bioinformatics* **2009**, *25*, 1754–1760. [CrossRef] [PubMed]
21. Broad Institute. Picard Toolkit. *GitHub Repository*. 2019. Available online: <https://broadinstitute.github.io/picard/> (accessed on 12 December 2021).
22. 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*, 491–498. [CrossRef]
23. McKenna, A.; Hanna, M.; Banks, E.; Sivachenko, A.; Cibulskis, K.; Kernysky, A.; Garimella, K.; Altshuler, D.; Gabriel, S.; Daly, M.; et al. The Genome Analysis Toolkit: A MapReduce Framework for Analyzing next-Generation DNA Sequencing Data. *Genome Res.* **2010**, *20*, 1297–1303. [CrossRef] [PubMed]
24. Poplin, R.; Ruano-Rubio, V.; DePristo, M.A.; Fennell, T.J.; Carneiro, M.O.; Van der Auwera, G.A.; Kling, D.E.; Gauthier, L.D.; Levy-Moonshine, A.; Roazen, D.; et al. Scaling Accurate Genetic Variant Discovery to Tens of Thousands of Samples. *bioRxiv* **2017**, 201178. [CrossRef]
25. Van der Auwera, G.A.; Carneiro, M.O.; Hartl, C.; Poplin, R.; del Angel, G.; Levy-Moonshine, A.; Jordan, T.; Shakir, K.; Roazen, D.; Thibault, J.; et al. From FastQ Data to High-Confidence Variant Calls: The Genome Analysis Toolkit Best Practices Pipeline. *Curr. Protoc. Bioinform.* **2013**, *43*, 11.10.1–11.10.33. [CrossRef]
26. Li, H.; Handsaker, B.; Wysoker, A.; Fennell, T.; Ruan, J.; Homer, N.; Marth, G.; Abecasis, G.; Durbin, R. 1000 Genome Project Data Processing Subgroup The Sequence Alignment/Map Format and SAMtools. *Bioinformatics* **2009**, *25*, 2078–2079. [CrossRef] [PubMed]
27. Botstein, D.; White, R.L.; Skolnick, M.; Davis, R.W. Construction of a Genetic Linkage Map in Man Using Restriction Fragment Length Polymorphisms. *Am. J. Hum. Genet.* **1980**, *32*, 314–331. [PubMed]
28. Hwang, J.; Li, J.; Liu, W.-Y.; An, S.-J.; Cho, H.; Her, N.H.; Yeam, I.; Kim, D.; Kang, B.-C. Double Mutations in EIF4E and EIF504E Confer Recessive Resistance to Chilli Veinal Mottle Virus in Pepper. *Mol. Cells* **2009**, *27*, 329–336. [CrossRef] [PubMed]
29. Perrier, X.; Flori, A.; Bonnot, F. Data analysis methods. In *Genetic Diversity of Cultivated Tropical Plants*, 1st ed.; Hamon, P., Seguin, M., Perrier, X., Glaszmann, J.C., Eds.; CRC Press: Boca Raton, FL, USA, 2003; pp. 43–76.
30. Sayers, E.W.; Beck, J.; Bolton, E.E.; Bourexis, D.; Brister, J.R.; Canese, K.; Comeau, D.C.; Funk, K.; Kim, S.; Klimke, W.; et al. Database Resources of the National Center for Biotechnology Information. *Nucleic Acids Res.* **2021**, *49*, D10–D17. [CrossRef]
31. Blast2GO: A Universal Tool for Annotation, Visualization and Analysis in Functional Genomics Research. *Bioinformatics* **2005**, *21*, 3674–3676. [CrossRef]
32. Shirayev, S.A.; Papadopoulos, J.S.; Schaffer, A.A.; Agarwala, R. Improved BLAST Searches Using Longer Words for Protein Seeding. *Bioinformatics* **2007**, *23*, 2949–2951. [CrossRef]
33. Ye, J.; Zhang, Y.; Cui, H.; Liu, J.; Wu, Y.; Cheng, Y.; Xu, H.; Huang, X.; Li, S.; Zhou, A.; et al. WEGO 2.0: A Web Tool for Analyzing and Plotting GO Annotations, 2018 Update. *Nucleic Acids Res.* **2018**, *46*, W71–W75. [CrossRef]
34. Reif, J.C.; Zhang, P.; Dreisigacker, S.; Warburton, M.L.; van Ginkel, M.; Hoisington, D.; Bohn, M.; Melchinger, A.E. Wheat Genetic Diversity Trends during Domestication and Breeding. *Theor. Appl. Genet.* **2005**, *110*, 859–864. [CrossRef] [PubMed]
35. The International Peach Genome Initiative; Verde, I.; Abbott, A. The High-Quality Draft Genome of Peach (*Prunus Persica*) Identifies Unique Patterns of Genetic Diversity, Domestication and Genome Evolution. *Nat. Genet.* **2013**, *45*, 487–494. [CrossRef] [PubMed]
36. Flint-Garcia, S.A. Genetics and Consequences of Crop Domestication. *J. Agric. Food Chem.* **2013**, *61*, 8267–8276. [CrossRef] [PubMed]
37. Fu, Y.-B.; Somers, D.J. Genome-Wide Reduction of Genetic Diversity in Wheat Breeding. *Crop Sci.* **2009**, *49*, 161–168. [CrossRef]
38. Park, S.; Kumar, P.; Shi, A.; Mou, B. Population Genetics and Genome-wide Association Studies Provide Insights into the Influence of Selective Breeding on Genetic Variation in Lettuce. *Plant Genome* **2021**, *14*, e20086. [CrossRef]
39. Ha, J.; Satyawatan, D.; Jeong, H.; Lee, E.; Cho, K.-H.; Kim, M.Y.; Lee, S.-H. A Near-Complete Genome Sequence of Mungbean (*Vigna Radiata* L.) Provides Key Insights into the Modern Breeding Program. *Plant Genome* **2021**, *14*, e20121. [CrossRef]
40. De Wilde, B.; Lefever, S.; Dong, W.; Dunne, J.; Husain, S.; Derveaux, S.; Hellemans, J.; Vandesompele, J. Target Enrichment Using Parallel Nanoliter Quantitative PCR Amplification. *BMC Genom.* **2014**, *15*, 184. [CrossRef]
41. Hollants, S.; Redeker, E.J.W.; Matthijs, G. Microfluidic Amplification as a Tool for Massive Parallel Sequencing of the Familial Hypercholesterolemia Genes. *Clin. Chem.* **2012**, *58*, 717–724. [CrossRef] [PubMed]