Article

Varietal Identification of Open-Pollinated Onion Cultivars Using a Nanofluidic Array of Single Nucleotide Polymorphism (SNP) Markers

Ji-Hee Lee [1,†], **Arif Hasan Khan Robin** [2,3,†] iD, **Sathishkumar Natarajan** [2], **Hee-Jeong Jung** [2] and **Ill-Sup Nou** [1,2,*]

1   Center for Horticultural Seed Development of GSP, Jeonnam, Suncheon 57922, Korea; jihee0830@scnu.ac.kr
2   Department of Horticulture, Sunchon National University, Suncheon 57922, Korea;
    gpb21bau@gmail.com (A.H.K.R.); sathisbioinfo@gmail.com (S.N.); my-656@hanmail.net (H.-J.J.)
3   Department of Genetics and Plant Breeding, Bangladesh Agricultural University,
    Mymensingh 2202, Bangladesh
*   Correspondence: nis@sunchon.ac.kr; Tel.: +82-61-750-3249
†   These authors contributed equally to this work.

check for updates

Abstract: Onions (*Allium cepa* L.) are a medicinally and economically important vegetable species rich in sulphur compounds, polyphenols, and antioxidants. In Korea, most of the onion cultivars are of the open-pollinated, heterozygous, short duration, and early spring type, which are generally harvested in April. Precise varietal identification is crucially important to warrant the authenticity of supreme onion genotypes, which aid in affirming the genetic identity of breeding materials at both the breeders and farmers levels. A set of markers identified from the double-digest restriction-site associated DNA sequencing (ddRAD-seq) database of *Allium cepa* L. and involving single nucleotide polymorphisms (SNPs) were deployed for genotyping deoxyribonucleic acid (DNA) samples extracted from seven genetically diverse onion cultivars collected in Korea and Japan. The validation process led us to designating 43 SNPs out of 48 that unequivocally identified all seven genotypes with high statistical validity ($p < 0.001$). Phylogenetic relationships and varietal identity among the cultivars were ascertained by Bayesian clustering and ordination analyses.·Two genotypes, Singsingball and Taegeukhwang of Korean origin, showed a greater genetic distance from the five other onion cultivars. The SNP markers deployed in this study effectively authenticated the DNA fingerprints of the early spring onion cultivars utilizing a high-throughput genotyping protocol. The method exploited in this study provides an efficient pathway of verifying genetic identity of onion genotypes for their quality control. The markers developed are highly useful in the management and conservation of elite onion breeding materials at the farmers' level.

Keywords: onion; nanofluidic array; varietal identification; SNP markers

1. Introduction

Onions are economically important culinary vegetables rich in sulphur compounds, polyphenol, and antioxidants, which all add medicinal value to them. Immature leaves and onion bulbs are widely used as vegetable. Onion salads and cut onion pieces are used in preparing various mixed-dishes, soups, and curries. They are often used as a spice and flavoring agent. They are rich in vitamin C, proteins and minerals, Ca and Fe [1]. *Allium cepa* is the most widely cultivated member of the genus *Allium* but related wild species also occur in central Asia [2,3]. The most related species is *A. vavilovii* [2,4,5]. The wild progenitors of *Allium cepa* are not well reported. Other notable *Allium*

species, such as *A. asarense*, *A. roylei*, *A. farctum,* and *A. fistulosum,* have quite distant phylogenetic relations with *A. cepa* based on sequences of inter-transcribed spacers (ITSs) of ribosomal DNA [5,6].

Plant breeders have adapted the domesticated form to a wide range of production conditions. The modern varieties were derived from heirloom varieties maintained within traditional farming systems. These varieties form open-pollinated, heterozygote, and heterogeneous populations [7]. Korea, Japan, USA, and many countries around the globe developed open-pollinated, short duration, and early spring type to summer type onion varieties. Open pollinated varieties may show a better yield stability over the years. In terms of the photosynthetic period requirement, the onion cultivars are generally categorized as long-day and short-day. Long-day onions begin sprouting in late spring or summer when day length reaches from 14 to 16 h while short day onions start sprouting in winter or early spring when day-length is in between 10 and 12 hours. In Korea, most of the onion cultivars are short day and they start sprouting in early spring [8].

Onion cultivars are very difficult to identify based on morphological traits such as foliage altitude or length, leaf glossiness, leaf color, leaf wax content, foliage cranking, bulb shape, bulb thickness etc. [9]. A number of simple sequence repeats (SSRs) have been developed over the last decade [10,11]. Despite this progress, no single and independent method has been developed so far that can identify onion genotypes. Single nucleotide polymorphisms (SNPs) are the most diverse and abundant type of polymorphism present in the plant genome [12,13]. SNP markers are more specific, reliable, and efficient compared with other DNA markers. Since SNPs are di-allelic in nature they provide a high level of consistency across different laboratories and they, in most cases, also offer very lower error rate in allele calling. In several recent studies these SNP markers have been utilized for varietal identification of different species, such as: cacao (*Theobroma cacao*) [14], grapevine (*Vitis vinifera*) [15], pummelo (*Citrus maxima*) [16], strawberry (*Fragaria* spp.) [17] and tea (*Camellia sinensis*) [18]. In a wide range of research and their applied fields, small set SNP-based DNA fingerprinting has great demand for economically and medicinally important horticultural crops like onions as well as many other perennial horticultural crops. These applications include, but are not limited to, identification of mislabeled accessions, parentage and sibship analysis for quality control in breeding and seed development programs, and characterization of farmer selections to support the production of high-value varieties for premium markets.

2. Materials and Methods

2.1. Plant Material

A total of seven onion cultivars collected from South Korea and Japan were used in this study for development of a nanofluidic genotyping system (Table 1). Four cultivars were collected from South Korea and three cultivars were collected in Japan. All of these cultivars were open-pollinated, short duration, and spring type.

Table 1. List of seven onion cultivars collected from Korea and Japan.

Cultivar Name	Code	Country
S&P2	P	Korea
Jelppareun	J	Korea
Singsingball	S	Korea
Taegeukhwang	T	Korea
Hamaemi	H	Japan
Gwigeum	G	Japan
Marushino	M	Japan

2.2. Genomic DNA Extraction

Genomic DNA was extracted from leaves of each sample using the DNeasy Plant Mini Kit (Qiagen, Wilmington, DE, USA). The concentration and purity of DNA was determined in a NanoDrop Spectrophotometer (Thermo Scientific, Wilmington, DE, USA).

2.3. Mining and Validation of Putative SNPs

Putative SNPs were selected from the restriction site-associated DNA sequencing (RAD-seq) of the Korean short-day onion inbred lines. In the double-digest restriction-site associated DNA sequencing (ddRAD-seq) procedure, genomic DNA at the first step is double-digested with two restriction enzymes, PstI and MspI (fast digest restriction enzymes; Thermo Fisher Scientific, Waltham, MA, USA). The restriction enzyme digested genomic DNA is then ligated to the Liga Fast Rapid DNA Ligation System (Promega, Madison, WI, USA). After purification and dilution of the DNA, a PCR protocol was used for the amplification. The amplicons were separated into DNA fragments from 300 to 900 bp using a BluePippin 1.5% agarose cassette (Sage Science, Beverly, MA, USA). The fragments were sequenced and SNP positions were selected based on the variable genotypic differences in single nucleotides. DNA fragments bearing more than 60 nucleotides coupled with at least three diallelic SNP positions were selected as SNP markers [19].

A nanofluidic genotyping system was deployed for evaluating and validating the selected putative SNP markers for their suitability in identifying heterozygous onion cultivars. For designing primers for a SNPtype^{TM} genotyping panel, 48 putative sequences (Table 2) were submitted to the Assay Design Group at Fluidigm Corporation (South San Francisco, CA, USA). Competitive allele-specific PCR and bi-allelic scoring of SNPs at the specific loci (KBioscience Ltd, Hoddesdon, UK) were the basis of the Fluidigm based assay. The genotyping kit developed by Fluidigm, SNPtype^{TM}, was used for this assay following the manufacturer's instructions [20,21]. Specific target amplifications using Fluidigm primers were carried out with a view to enrich the required SNP sequences. An integrated fluidic circuit developed by the Fluidigm Corporation was used to perform genotyping in a nanofluidic 48.48 Dynamic Array^{TM} IFC (integrated fluidic circuits) [22]. The IFC chip assembles PCR reactions automatically and enables testing of up to 48 samples involving 48 SNP markers simultaneously. An EP1^{TM} imager integrated with 48.48 IFC was used to acquire endpoint fluorescent images. The Fluidigm Genotyping Analysis Software was used for data analysis [23].

Table 2. The flanking sequences and single nucleotide polymorphisms (SNPs) of the 48 polymorphic markers.

SNP IDs	SNPs with Flanking Sequences
O-SNP001	TATGTTTATGAAAATAAATTCTGAAGCATAATTCATATTATATTTTACAAGTTTTTTATATGAGT GTTCAGTTAAGCGCACAGGTTACATCATTAGTTGG[A/C]TTGTTTATACAATTGGACAAGTTC ATTATGTAGGTAAACTTGTACAGTTTATAGAAAATGCTTTAATCTTCAATTATCACTATAAATT TACAGCTTGTGA
O-SNP002	AATTCACTGTCAAATCCCTTGGTTTAGAACAAGGAAATATGCTACGTTAAATAGCTATAGCATC GTCGACAGTTTATTCAT[A/G]TTATATAGCTAAAGAGAATTGTCAGAAGAAAGCAATTATTACA TGTTGAGGGCAAAAACAGTAATTTCTTACCATTTCCAACTTCTTCATGATGTTCTAA
O-SNP003	AATTTCTCCTACAGTTTTTCCTTCACAAATTCAAGCATGCATTTTTCGAATTACAAATACTAACA CTATCCAGCTGTATAGCATGAAGGACAATATGAGC[A/G]GTGGTTTAAACTTCGTTTTTAACAT ACCTTTGAAGGTATGCAATGCCAAGATTTGCAAGGGAATCACAGTTATCTGGAGCAAGATC CAACACGGATGATA
O-SNP004	AATTCGAACCATGTGGGTACTCTATGAATGGAAT[T/C]GTCGACTCTGCACTGTCTACTGTACA CGTTACTCCTGAAGATGGTTTTAGCTATGCTAGCTACGAAGCAATGGGGTTCAACCCTGAAA GAATCGATTTTG
O-SNP005	GCAGTTGAGTTGACGAACTAGATCCGATCGCCTCCACATATGTATGTACGATGATTAACATTGC TTGTGATTTGGTC[A/T]GTCTCGGGATCAGATGACCGATAGGCCAACATTTTCTTCCAAAAGCA TAGGCTCGAAAGGGTTAGGTTGGCTCATCTTTGACGACTTGAGTCTTATTATG

Table 2. *Cont.*

SNP IDs	SNPs with Flanking Sequences
O-SNP006	CACCTCCTTATCGTTGGCCTGATTGATATACAAAAAACAGTATACTTCAACATAACTCAAATGT AGCCTCTTGACTCTGGGATACCCTTAGCAACAAAAC[A/G]AATAACTCTTGCCCAATGTGTTC ATTTATATGATACGATAACTGC
O-SNP007	GCAAAATATTTCCTGTAAACATAGAGAATTGTGGAGTCATCAGAATTCATGTTTGTGTTA[T/G] AAGTCATTTATTGTTTTATGTGGAGTTCATTTGATTTGGTCCAATTCAGACTTCTTATGCATCTTT TCCCCTTTAACGTGCAACCAAGAGCATGTGATTT
O-SNP008	ACATATACATTTTCATGACCATTCATCAAGCTTCTGTACAAAGACGTAATAAGAATTCACCTCA ACAACATTAAATTAATAAAAGCACTTTAATGACAGC[A/C]CTCGTCGAGCTTTTATACACACA AATGAACATGGCAACAATCAGAATCTATGGCAATCCAAACCAACCGAAATAAATAAATAAAT AAACTAAAAAAAGAG
O-SNP009	TCAGTTCAATTTCTGCTCATCAACCATTTACGTGCAAATTTTTCACAAATACTTCATTTATGTG CAGCAGCACCCTCGGCCAACAGCTACCACCACGTCT[A/T]CACATTCCCTCCAACTGACAAT CTACAAGAATCGCGAATT
O-SNP010	GCAGTATAATGTATTGGACTAAATATATCATTTCAAGTTAAAATATTATATTGATTAAACTACAG TATTCTGACGGGCTTGATTGAACA[A/G]TTGCTTTTTGAATTAATTGCACATTATGGAAGTATTT GGCAAGTCATTTTCCGGGAAAGCAACAAGGAGTGTCAGCCTCTTTCAATTGAATTAGATGTAT
O-SNP011	GAACCAAAAAGGAATTACCGCATGTCCAGAATTCCATCGGAATTTTAAACCATGGAAACATC TCTCCCACTTAAATAG[T/G]TATACGAGTACATACCTATCAATGGACTTTTGCGGAGAACGCAT AAAAATGAACAGTAAAGTAAACGTAAGTACCTTTTTTTCTTTTTGTATCTCAGATA
O-SNP012	AATTCCAGCGTTATGCACAGAATTAACTAATTATGAGTGTTCTTATGTGGTTGATGACGATTCA GAATGAGTGGTATA[T/G]CTTACACTCATGTCAAGGTGGTTGTATGTGATCATTTAATGGGATC ATTTGTTAATATGAAGAACTAAAGTTGGCAATTAAGGAAGATTCATGAATTTAT
O-SNP013	CTCAAAATGCTGATGTACAACTTTACCAGCATCAGACAGTTCATTTCGGTTAAATTAGCTAAAT TGACCACTAAACAACATGAACTCTTTACTTCTTCAT[C/T]GCTCAACATAATTTTCAAATTAAA ATGGATGGGTGTGCATGTATGAACAATAACGAAACAAAACATCTGC
O-SNP014	GCAGCTTCACATCTCATTCCTTCTGTTATTATGTAC[A/T]TGCTTCCTGACTTCTTCCATTCATG CCCACAACTTCACATTTCAACTGTTGTTACGTGTATGCCAGCAGCTCATTCTCTTTCKTGCATG CTGAATTTTGAAAAGGTTTTAAAACTGAATCTTGAATT
O-SNP015	GCAGAAAACGAAGGCCTTGATTATGAAACTGTATGTTATTGTAAAATTACTTTGAAATCTAACT TTCTATGAA[T/G]TCATGGGCTTAATCTGATTGAAATTCCTGCACTGTTTTTTAGGCTATGGGA AGAATTTTTATCTAAAGGGTTTTAGGGAAGAATT
O-SNP016	GCAGCCACTGGTGATGTTGGTYAGTGGTCTCGGTTCAATAAGCTTGTCCTTGGTGCTTGTATT TCTGATATTGAAGATGTTGGTAGTGGTGAT[T/G]TGGTAGTTRGGCTTCATGAAGTGCATGGA ATTTATTAAAATAAATATTAAATATTTGAAAAATT
O-SNP017	TGGTGCCAAGAGCACGGAATTCAGATTTACAGACTTAAGATATTCATATGCACAAGATTAACA CGCAAATACTCTATATACATAATGGAATCGGTATAGT[C/T]TGCAAAGAAACAAAACCTACTAT TGTTAGAGTATTAAATTTCCATGAGAAACAGATTAACCGAGTATATGAGCAGATGGAATTATTC ATTATAGTTTGCA
O-SNP018	GCAGTTGGTTGTAGTTGTGGAGTCTAATTTAGTAACACCCTTTGTTAAAATAACCACTGAATC [A/T]TATTAGTGAATGTTGGCGGTACAAGGAATCTTTTTTTCCCCTCCAAAATCAACTAATTTTA TTGGTGAATTTTACGGGGAACACCAGTGCTAACAAAATT
O-SNP019	GCAGCTTCAGCTTCAATAGATGGTAATATATCGGATGCCAACCCCGAATTT[C/G]TGTTACATG TTTCGTTTTGCTCATTGGGTTCCACTTCAATGGAAGCTATTATCAGATCTGATGTTGATTTCGG AACTTCAGAACATCTTTTGTTTAGTTC
O-SNP020	GCAGTCATAACAGCAGGTAAAGTAATCTTTCTTTTTTATACTTGTCATTTCCGCACTAAAGTG TAAGGGCTTGT[C/T]GTTGTGCCATTACCATATTAAGGAGCTGCTTGTTTGATGAAAATATAAG TAGGCTTGGGTTATCATCGTGGGAAGATTAGTATATGGACTGATAGAGTAA
O-SNP021	ACTCCTGGTCCCTCAAAAACCCTAAACCCAATCCCAAAATCTCAACTTCATCAAATGAACCTC AAACTCATGACATCAATTATTCTCAAATGCTCGACCA[A/G]TTCTTTTACGAATGCGAGCACCG GCCTGATTACCGTCACACTCCCGAGGTTGAGAAAATTTTGAATT
O-SNP022	TGAATTATAGCAGCGAGATATATGATATATATTAATTGAAAAGAAAACATCTGCATTATACAAA CAAACTAAAATGTATAGGCAAATACAATATGTGGGG[T/G]GTGTGTGTGTAAACTGGACTGCA ATTTTGAAGACGTTAAAGTGAGAATT

Table 2. *Cont.*

SNP IDs	SNPs with Flanking Sequences
O-SNP023	ACCTGCTTCAGAAAAGAAGGGATTAGGTTATGTTAAGAATTCCAAAGAAGTGGAAATCTGAG TTGTAAATTTGTGAGACAAGTTTATTATTGATTAGTCT[A/C]TGTTTTGGTATACTAGACGTATAA CATATTCTGTACTAACCTACTCAACCTTTCATTTCAAATTTTCTAAAACAGGGTACAATGGCTG ATGAATTAGATC
O-SNP024	CGATCTGCAGAACATAGCATGTCTGATCATACATCAGAACAAGATTT[A/G]ATAAACTCTATTA TCGTATCTGCAACTCCTGGTACTTCTAATGAATGCATTGCAATAGCTTGTCAATCTAATGGTAA ATTGTTGCAATACCATCTCACTC
O-SNP025	AAAATGAGTTAAAGGCAAAAAATAGTTGTATTAAACAAGACATAACATTATGCTTAATAAGCT TATTTCCCGACATCATATGCACATGGAGATCAATTGT[T/G]TTGTAAAACTGTTCACTAGAAAT CATGTCATGTGAATGGAAAAGACATGCTCGTCTAAGTTGTAGTAGATGGGATACACTGC
O-SNP026	AATTCACTTTCAATCCTAATTTTGACATGAACTTC[C/T]CATTTATAAGTATGGGTGCCTGAAGG TTTATTACAGATTACTGATGAAAGAACTGTTAAATTCTTGGAATTATCCATTTGATTTATCGCAG CATGTTACA
O-SNP027	AATTCAGCTGGATCAGCTAATATCAATCTATGAACAATTCGATAGTACTGATCTTGTTGACGAT ACATGCTATACATAACAATTTTGCAC[A/T]GCAGTATCTTTTATCCAACTACAAATTATGTTTGT GTAATAAACATGATAATTTTTTGGTATGTGGTAAATCTTTAAATAAAATTAGCATGAGAAATGT
O-SNP028	GCAGTTCATTCATCACATATCTAAGTCAGAAAACACCTTCATATTTATCTATATGTAACTCGCA [T/G]CATAGTATTTTCTTCTTTAAGGTTTTTGAGTGCATCAGATTTGGACGTGCAATCTTTGAAC TTACAATATAGAACTTAGTTTTTGACTACCTCAGATTCG
O-SNP029	TCGTTTTGATACGTACCTTTATTCCTTTTTGACATAGAACAGAGCTTAATAAAGATGCCACTTTA GAACGATGTCCAACCAATAAAACATCTGTAGTGTT[C/T]TTGGGAATATTATTCACAACCACTGC
O-SNP030	GCAGTTCCGACCCCTCCGAATCCAATTCGCCCTTTTCGGATCGGATCCTCTACATCAAAAT[C/T]G CCAACAGGTTGCTTTTGGTTGTTTCTCCTGCACGCATCGAGCTCTGGAGCTCATCTCAGGTAACA GGTGAAATTACTTATTGACCCTTGTTCGAGCTCG
O-SNP031	CCACAGAAAATGTGCCATGTAGATTGGATTACATTGATTGACATAGTGGACAGAGGGCTGGACA GATGGAAAACACTGCTTGACCAAAAATTGGAAGACT[A/G]CTAGTATGTTTCCATGTGCAAGTTT GTGCTCTTTATGTATATAGGTGCCTAAAACTAAAGAGTGTCATGTGAAGGGTAGCTGC
O-SNP032	TAATCCCAAAAAAAAAGCAACTTTATAATTAAACCTCGTAAACGAAGCAATACAAAGGAAACCC TAACCTATTAATTCATTTAACTGATGCAGCTCCATT[A/G]ACGATTTGCGAGGGATTCTGAGAAGC ATCTTCTGC
O-SNP033	TCTGATCCTTGAATTGCTTTACAGCATGAAGCTGAAGGAGGAACAATTGGACCTGGCTTCTGAAC GTACGAAGCACAATTCTTTGTCAGTCCATTAAAAT[C/G]GCCTTTGCATGGACCAGAAGGTGACA TATCATGATCCATGTAATCTACCGCTGCTGC
O-SNP034	AATTCTGACTTTAAGCATCATTGAACAAATTATAATCCTTGATTAGCATATAACTTCTCATTATTTTA TACATGTTTATTAC[A/G]GAGTCGGTAGGGAGATATGGTCAGATTAGTATGCATGAACTCATCGAT CGGTTTGATAAGGATATGGGATTACTTTATTCATTTGAATTATTTTATGATT
O-SNP035	GCAGCCCCTCCAACACTCCTATAACCACAGTCAACATAGCCCCATCTCCATAC[A/C]GTCCCCA TCATATCCACCCACATCCTCTTCCAAACCCACAGCCCATCAAATACACTACTGTTGCCTTACTGCC TCAAACACCATCAACACAGTGCTGTGA
O-SNP036	AATTCTATGACCATTATATGCAAGAATACATCAGCAAAATTGCGATCGATCGCTTGCCGGAATCA A[A/G]ACCTCTATGGAACATTCACATAATCAAATATCCTACAAAACAGGCAGCGGGTACACTTAT ATTCAAACTTCATCATTCTTTAGGAGATGGTTTTTCTTTA
O-SNP037	CCTATTCAACGGCACTTCACCTAACAACTTCCTCTGATCACATACCATACACCTTGATCAACCTCA TTCCACCATCCACCTTGATCAACAGCAGCTACAA[C/T]CACGAGTTCTGTGCACATTATATCATCA ACATCATTTACCTAAACCCAAACTTCATCAAACTGTCACACTGAATT
O-SNP038	AATTCTTGTTATTTCGATACACCATATTTAACCATTTTCAGATGAATTGATCTACTTGACTTAATCTT AACTATGA[A/T]GTGGTTAATGTATCTACAGGTGGCCATATCCTTTCCTTGAACTCTCGACTCCTTG GGCTCCAGTTTGGTATGGCAATGATCAAGTCTTATTTGATTCCGA
O-SNP039	ATATTAAGCATTTACCTCCTCTTTGTGAATCCTTCCATGGAGGGAAAGATGATTTGATCGTGGAAA ATCATTGCTTATTATCTTCAATTTACACTGATGT[A/G]TTTCTCTAGCGGCCTTGAAAAACTGC
O-SNP040	CCTTACCATCCCGAATTCTGAAAATCCAAATTGATTAAACCATCGAAACCCTAATTTCAAAAAAT ATGAAGAAGTACGGTCTTAATCTTCGTATACCACC[A/T]AATCAGCAAAAGAAACAAGCTTCT TCTTCTTCACAACCTTTATCAATCCGTCCCCCTCCTTCGATTTTTAAAGAAGAAGACGATGATGAG AATGCAGTGG

Table 2. *Cont.*

SNP IDs	SNPs with Flanking Sequences
O-SNP041	GCAGTGAGATAACAAAACAAGTTGTAGCTAAATGATGCAAAAATGACACGCACTTACCA[T/G]T TTGCTTAACTCAGGTGGAACTTTCAAATTTTCGATGAAGTCAGCAAAGCTAACGGCCTTCTGCCA CCAGCCACTGTCTGACTTACAGCCGTCACACGCC
O-SNP042	GCAGTTATCCAGCTCAAGCATAAAATAAAACTCACCCAAATACAATACTTATTTAAGCATTGAGC TCTATGTGTTTACT[A/G]ATACGTCCATACGAGACTCAACATTAATGTAGCTCAATTCACTGACATA TTGACCAAAAAACCACGACTTTACGTGATTTCTTCCCATCCATCAATCTCC
O-SNP043	TTTCTCTGACAGCGTCCAGGCTCGACGAATTCGCATATCAATCTTGTCTCACTTTTTCTTAATGAAT TAAGCTCGCCGCATAGTGCTGGACTTGATATTG[C/T]TGTCAGTAATTTTGTTTAAAAGTCTGTTGT CTCTGCTTTTATCAACTGC
O-SNP044	GTTTTGTTTATATCTAATGCCTTAATTTCTTAAGCCAAGGCTTGACGCCAACAGGGATATTTCATAG CTTAAGAAAAAGAATTTGGTTCCTCAGTGGTGA[A/G]AACTCGAGAAAGAAAGATCCTATCAGA AGATTAAAATTCATAAAGATGACTTAAAACTACTAGAATAGTTTAGAATT
O-SNP045	AAAACAGTCCAGTAATTCTATATTATAAAGCTCCGGTCCTTATTTCCAGGTCTAAGCATTACACAA CATTTTAAAGTTGAAACCTCAAGCGATTTTTAAC[A/G]ACTTTACTCAATGATGCAGTGGCACTTA AACTGC
O-SNP046	AGCAAGTACACGGACTGAATCTAGCTATCCTATACATAGGACTGTACACACTCGCATTAGGAGAA GGTTTTATGCGAGCGTGTCTGGCATCATTTGGAGG[A/C]GACCAATTTGACAGCGATGATCAAAT CGAGCTGAGACAGCAGTCTAGTTTTTTCAACTGGTTCACTCTCGGTATATCTGC
O-SNP047	TTAAAAAAAAAATACTTGTAGAAACCACCATAAAATTATGCCTCGTAGGTTTTGAATCTGACTAAT GATCTACTACATGCATGTATCATGAAGATGAGACA[C/T]AAAATGAAATATTAAGCACTTCTCAA GTAAGCTTTAGAAGAGTACTGC
O-SNP048	ATGTTTCAACTTTCAAGAGTAAAGAAACAAATGTTTCAATTACAGGTTGTCGCTAAATTCATGTCT CAAGTGCTACGCACTTCATATACATTCAGCCCGT[C/T]GCAAATAATAATAACTTCTGGCTCTACT GCTGCAG

2.4. Data Analysis

Minor allele frequency (MAF, the frequency of the least frequent allele in a given locus and in a given population), expected heterozygosity, Shannon's information index, and the inbreeding coefficient were the key statistics calculated to describe the patterns of genetic variation within the collection of seven onion varieties. The statistical computational software GenAlEx 6.5 was used for enumerating descriptive statistics [24]. A pairwise multilocus matching among individual samples was used for identifying genotypes using the same software. The identical SNP genotypes obtained from different cultivars were considered to originate from the same parental combinations. It was presumed that two different individuals, of the same cultivar or of two different cultivars, may share an identical locus since the genotypes are heterozygous. Therefore, statistical rigor was assessed using probability of identity (PID). A multiloci-PID value was obtained from the product of PID values of individual loci, considering that each individual locus is independent. Further, a separate multivariate statistical approach, deploying distance-based analysis, was conducted for assessing inter-relations among the individual genotypes as compared to a reference genotype retrieved from the international genebanks. The DISTANCE procedure protocol from the GenAlEx 6.5 software was used to calculate genetic distances [24]. On the basis of the pairwise Nei's distance matrix, a principal coordinate analysis (PCoA) was conducted using the same software. Standardization of the distances and covariances was conducted. The genetic inter-relations were further examined after a cluster analysis, as a complement of PCoA, and a dendrogram was drawn where the neighbor-joining algorithm from the MEGA 6.0 program was used [25]. As previously used in SNP data, the STRUCTURE software with its default algorithm was used for a model-based clustering [26].

3. Results

3.1. SNP Discovery

A total of 12,719 scaffolds of Korean short-day onion inbred lines were obtained. A comparatively smaller number of 558 scaffolds were selected to identify 1904 SNPs. The characteristics of the SNPs were 1326 transitions, including 660 A/G and 666 C/T, and 578 trans-versions, including 266 A/T, 141 A/C, 110 T/G, and 61 C/G. In this study, we selected 48 SNPs for the genotyping test (Table 2).

3.2. Frequency of SNP Markers and Descriptive Statistics

A total of 43 SNP markers were genotyped successfully out of the 48 selected SNPs. The remaining five markers were not genotyped, probably because of the presence of complexity in sequences of onion cultivars or maybe due to presence of polymorphisms within the flanking sequences. All 43 markers were found polymorphic across the seven onion cultivars (Supplementary Information Figure S1).

Allele frequency of the markers ranged between 0.504 and 0.857, whereas the mean allele frequency was 0.652 (Table 3, Supplementary Information Table S1). The mean Shannon's information index was 0.727, but that value ranged from 0.233 to 0.958 (Table 3). The expected heterozygosity of the SNP markers ranged between 0.146 and 0.571 with an average value of 0.444 (Table 3).

Table 3. Allele frequency, Shannon's information index and expected heterozygosity of 43 SNPs.

SNP ID	Allele Frequency	Shannon's Information Index	Expected Heterozygosity
O-SNP001	0.786	0.656	0.413
O-SNP002	0.656	0.835	0.494
O-SNP004	0.54	0.822	0.5
O-SNP005	0.657	0.942	0.568
O-SNP006	0.732	0.586	0.345
O-SNP007	0.772	0.56	0.334
O-SNP008	0.527	0.691	0.431
O-SNP009	0.571	0.843	0.521
O-SNP010	0.594	0.892	0.56
O-SNP011	0.567	0.864	0.516
O-SNP012	0.728	0.796	0.507
O-SNP013	0.518	0.871	0.536
O-SNP014	0.629	0.921	0.571
O-SNP015	0.536	0.883	0.542
O-SNP017	0.603	0.69	0.42
O-SNP018	0.759	0.72	0.435
O-SNP019	0.638	0.927	0.549
O-SNP020	0.701	0.608	0.371
O-SNP021	0.813	0.709	0.434
O-SNP022	0.634	0.9	0.538
O-SNP023	0.665	0.918	0.567
O-SNP024	0.754	0.233	0.146
O-SNP025	0.661	0.907	0.555
O-SNP026	0.795	0.712	0.43
O-SNP028	0.504	0.958	0.571
O-SNP029	0.786	0.554	0.33
O-SNP030	0.674	0.401	0.244
O-SNP031	0.603	0.694	0.444
O-SNP032	0.563	0.783	0.489
O-SNP033	0.692	0.598	0.36
O-SNP034	0.652	0.779	0.491
O-SNP035	0.763	0.6	0.357
O-SNP036	0.54	0.493	0.313
O-SNP037	0.522	0.723	0.435

Table 3. *Cont.*

SNP ID	Allele Frequency	Shannon's Information Index	Expected Heterozygosity
O-SNP038	0.857	0.508	0.308
O-SNP039	0.746	0.85	0.511
O-SNP040	0.527	0.835	0.507
O-SNP041	0.79	0.745	0.471
O-SNP042	0.629	0.673	0.404
O-SNP044	0.589	0.736	0.441
O-SNP045	0.54	0.713	0.44
O-SNP046	0.531	0.744	0.461
O-SNP047	0.692	0.405	0.231
Mean	0.652	0.727	0.444

3.3. Cultivar Authenticity

The genotyping data were consistent in repeated tests, conducted independently for three times, with onion DNA from the seven cultivars. Identical SNP profiles were obtained from each of the cultivars for each of the three replications (Table 4). A pairwise comparison based on 43 markers revealed matching only for individual genotypes (Table 5). There was only a 0.000001 probability that two onion cultivars would produce the same genotype at the 43 SNP loci, as computed by the multilocus matching procedure implemented in GenAlex 6.5.40. Singsingball (S) and Taegeukhwang (T) had a high genetic identity between them, whereas these two cultivars had a smaller Nei's genetic distance (Tables 4 and 5).

Table 4. Pairwise population matrix of Nei's genetic distance.

S&P2	Jelppareun	Singsingball	Taegeukhwang	Hamaemi	Gwigeum	Marushino	
0.000							S&P2
0.143	0.000						Jelppareun
0.542	0.762	0.000					Singsingball
0.563	0.773	0.129	0.000				Taegeukhwang
0.082	0.125	0.598	0.614	0.000			Hamaemi
0.121	0.253	0.570	0.591	0.168	0.000		Gwigeum
0.139	0.181	0.553	0.565	0.125	0.106	0.000	Marushino

Table 5. Pairwise population matrix of Nei's genetic identity.

S&P2	Jelppareun	Singsingball	Taegeukhwang	Hamaemi	Gwigeum	Marushino	
1.000							S&P2
0.866	1.000						Jelppareun
0.582	0.467	1.000					Singsingball
0.570	0.462	0.879	1.000				Taegeukhwang
0.922	0.882	0.550	0.541	1.000			Hamaemi
0.886	0.777	0.565	0.554	0.845	1.000		Gwigeum
0.870	0.834	0.575	0.569	0.882	0.900	1.000	Marushino

3.4. Genetic Relationship among Onion Varieties

Principal coordinate analysis clustered seven onion cultivars into two major groups across the first coordinate (Figure 1). Coordinate 1 separated two Korean onion cultivars Singsingball (S) and Taegeukhwang (T) from the other five cultivars, with contrasting positive and negative coefficients, respectively. Coordinate 2, on the other hand, separated Jelppareun (J) from Gwigeum (G), with contrasting positive and negative coefficients, respectively (Figure 1). The other three cultivars, Hamaemi (H), Marushino (M), and S&P2, clustered in between Jelppareun (J) and Gwigeum (G). Similar to the PCoA in Figure 1, the neighbor-joining dendrogram also separated the two Korean onion cultivars Singsingball (S) and Taegeukhwang (T) from the rest of the onion cultivars (Figure 2).

These two cultivars, S and T, produced a distinct cluster whereas there was no definite pattern that can separate the other five cultivars.

Stratification clustering of the seven onion cultivars, computed by the program STRUCTURE, revealed that the two cultivars Singsingball (S) and Taegeukhwang (T) had similar multiloci when compared to the other five cultivars (Figure 3). This inferred clustering was compatible with clusters obtained from PCoA and the neighbor-joining tree. The cultivar Jelppareun had the most dissimilar genetic identity compared to Singsingball (S) and Taegeukhwang (T).
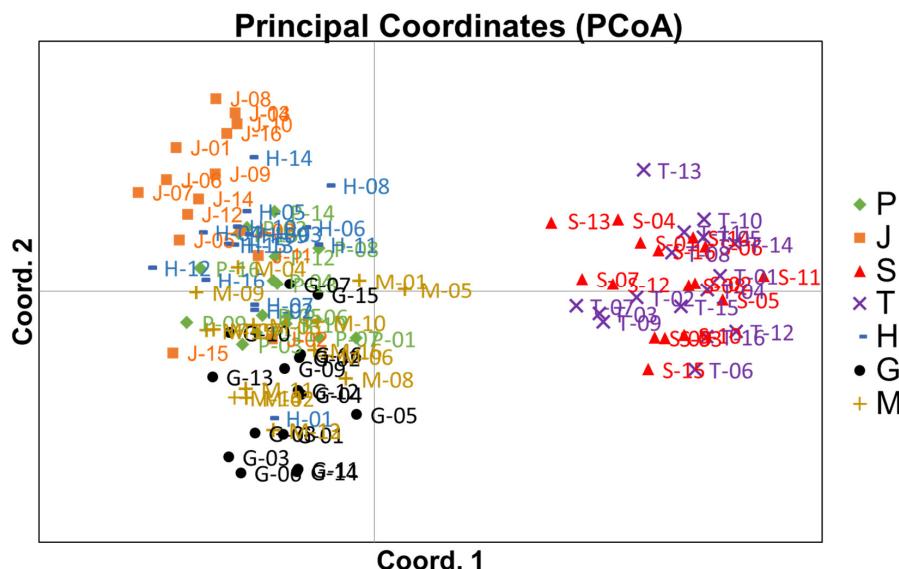
Figure 1. Principal coordinate analysis (PCoA) reveals associations between onion cultivars according to evolutionary relationships. P, J, S etc. are the cultivars name as defined in Table 1.
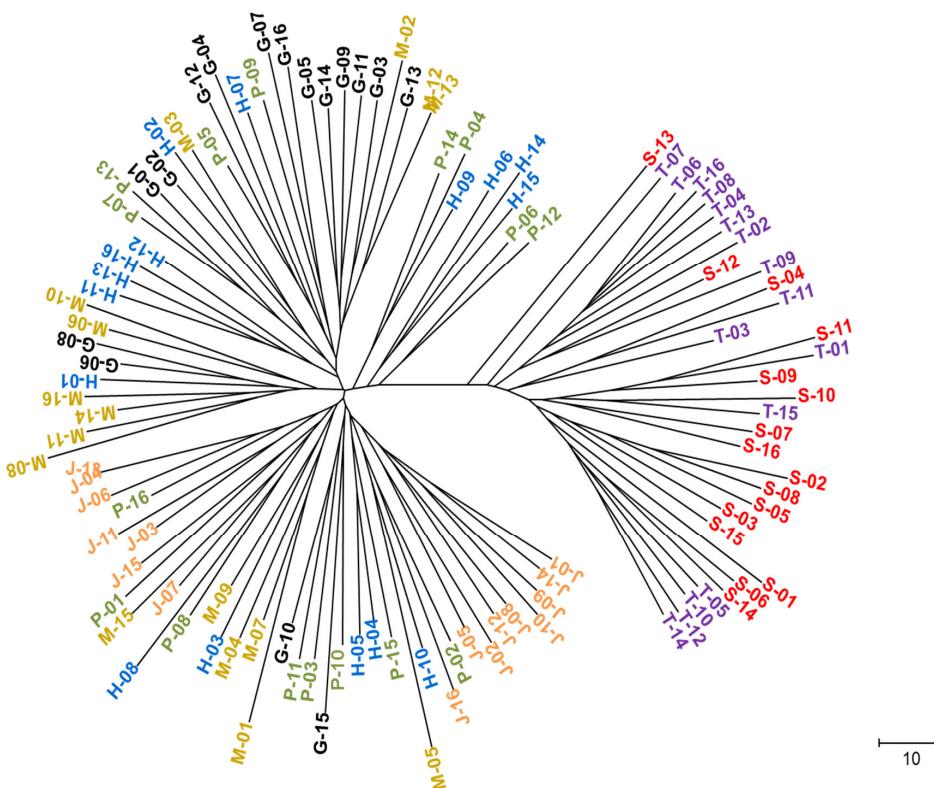
Figure 2. The evolutionary relationships of the cultivars according to the neighbor-joining method drawn using MEGA6. The branch lengths indicate the evolutionary distances within the phylogenetic tree.
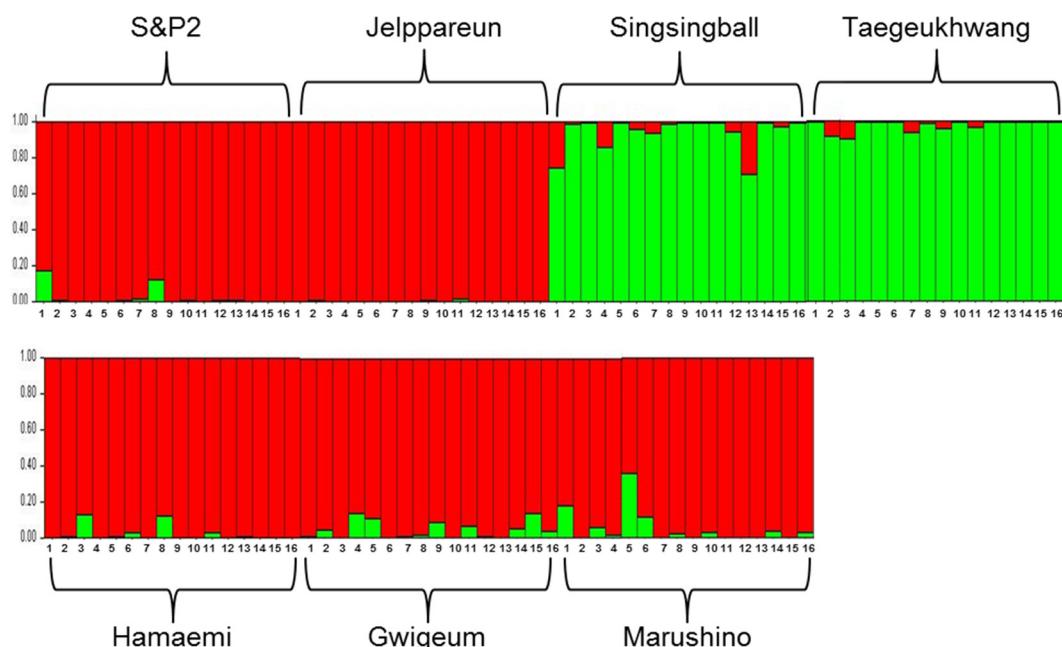
Figure 3. Inferred clusters in the onion varieties using the STRUCTURE software, where *K* is the potential number of genetic clusters that may exist in the overall sample of individuals. Each vertical line represents one individual multilocus genotype. Individuals with multiple colors have mixed genotypes from multiple clusters. Each color represents the most likely ancestry of the cluster from which the genotype or partial genotype was derived. Clusters of individuals are represented by colors.

4. Discussion

Onions, as vegetables, are becoming increasingly important at a global scale. To date, no efficient genetic method has been devised that can identify onion cultivars. Due to absence of sufficient throughput and adequacy in standardization of data, existing molecular technologies have limited use e.g., SSR marker-based DNA fingerprinting [10,11].

The present study exploited a small set of SNP based markers to develop a DNA fingerprinting, method with a view to distinguish onion genotypes. By utilizing this method a reasonably large number of onion leaf DNA samples can be handled to genotype them in a comparatively shorter time period. The markers and the methodology are highly reliable and the results are repeatable. The nanofluidic array of SNP markers is therefore suitable and applicable for the genotyping of open-pollinated onion cultivars. Our protocol, specific target amplification, most efficiently handled the possible drawbacks and maintained the quality and quantity of genotyping DNA materials (Figure 1). The specific target amplification protocol amplifies all loci of interest and thus increases the copies of those loci. This SNP protocol has recently been used in several crop and plant species including quinoa (Chenopodium quinoa) [27], Amaranthus (Amaranthus sp.) [28], Rosoideae [17], cacao (Theobroma cacao) [14], tea (*Camellia sinensis*) [18], longan (*Dimocarpus longan*) [29], coffee (Coffea spp.) [30], etc. This protocol thus can utilize a low concentration of DNA for the genotyping of the commercial onion cultivars, similar to the other plant species. One of the remarkable feature of the data was 100% genetic similarity in three independent samples, although the samples were extracted separately. Therefore, it could be argued that the nanofluidic system is highly efficient in DNA fingerprinting, with greater accuracy and higher reliability.

PID is the statistical parameter that determines the number of loci that are required to identify a distinct genotype with a required level of confidence. That means the PID value, reflecting the number of loci involved in a specific genotyping, determines the effectiveness of the genotyping. Since onion is an open-pollinated domesticated crop species and its progenitors are inbreeds, a stringent PID value

is essential in genotyping this species. In the present study, our results suggest that utilization of 43 SNP loci created an opportunity to sample an identical genotype with greater efficiency; 1 out of 100,000 from a random mating population. In this way, PID indicates its high statistical reliability and efficiency for onion cultivar identification.

5. Conclusions

In conclusion, the present piece of research conducted a SNP marker-based study to identify open-pollinated onion genotypes. This study was primarily developed with a set of 48 SNP markers for genotypic identification of open-pollinated onion cultivars but found 43 effective markers useful in a nanofluidic array. The makers and the methods created an opportunity of generating high quality SNP profiles based on fresh leaf-extracted DNA samples of onions. Taken together, these statistical tools, SNP markers and technologies opened up a new window of genotyping onion cultivars for unambiguous identification.

Supplementary Materials: The following are available online at http://www.mdpi.com/2073-4395/8/9/179/s1, Figure S1: Call map view from the dynamic array of single nucleotide polymorphism (SNP) fingerprints of genotyped onion varieties from the Fluidigm Genotyping Analysis Software. Table S1: Allele frequencies in 16 samples (*N* = 16) from each of the seven open-pollinated onion cultivars at the 43 different SNP markers. Allele/n designates the positions of the SNP under each marker, which are variable across the seven cultivars.

Author Contributions: I.-S.N. conceived and designed this study. J.H.L. and H.J.J. designed the experiments. J.-H.L. isolated the DNA and conducted experiments. S.N. conducted all statistical analysis and prepared the figures. A.H.K.R. organized the tables, wrote the manuscript and prepared all revisions based on reviewers' comments.

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

References

1. Cope, R.B. *Allium* species poisoning in dogs and cats. *Vet. Med.* **2005**, *100*, 562–566.
2. Fritsch, R.M.; Friesen, N.; Fritsch, R.M.; Friesen, N. Evolution, domestication and taxonomy. In *Allium Crop Science: Recent Advances*; CABI: New York, NY, USA, 2002; pp. 5–30.
3. Grubben, G.J.H.; Denton, O.A. *Plant Resources of Tropical Africa 2. Vegetables*; PROTA Foundation: Wageningen, The Netherlands, 2004.
4. Fritsch, R.M.; Matin, F.; Klaas, M. *Allium vavilovii* M. Popov et Vved. and a new Iranian species are the closest among the known relatives of the common onion *A. cepa* L. (*Alliaceae*). *Genet. Resour. Crop Evol.* **2001**, *48*, 401–408. [CrossRef]
5. Silva, M.L.R.B.; Magalhães, A.G.; Mergulhão, A.C.E.S.; Candeia, J.A.; Silva, M.C.L.; Lyra, M.C.C.P. Phylogenetic relationships of onion (*Allium cepa* L.) cultivars by sequencing of the its region. *J. Exp. Biol.* **2013**, *1*, 6.
6. Gurushidze, M.; Mashayekhi, S.; Blattner, F.R.; Friesen, N.; Fritsch, R.M. Phylogenetic relationships of wild and cultivated species of *Allium* section *Cepa* inferred by nuclear rDNA ITS sequence analysis. *Plant Syst. Evol.* **2007**, *269*, 259–269. [CrossRef]
7. Rivera, A.; Mallor, C.; Garcés-Claver, A.; García-Ulloa, A.; Pomar, F.; Silvar, C. Assessing the genetic diversity in onion (*Allium cepa* L.) landraces from northwest Spain and comparison with the European variability. *N. Z. J. Crop Hortic. Sci.* **2016**, *44*, 103–120. [CrossRef]
8. Lee, J.H.; Natarajan, S.; Biswas, M.K.; Shirasawa, K.; Isobe, S.; Kim, H.T.; Nou, I.S. SNP discovery of Korean short day onion inbred lines using double digest restriction site-associated DNA sequencing. *PLoS ONE* **2018**, *13*, e0201229. [CrossRef] [PubMed]
9. Ahmed, N.; Khan, S.H.; Afroza, B.; Hussain, K.; Qadri, S.; Nazir, G. Morphological characterization in onion (*Allium cepa* L.) for preparation and implementation of plant variety protection (PVP) legislation and distinctness, uniformity and stability (DUS) testing under temperate conditions of Kashmir. *Afr. J. Agric. Res.* **2013**, *8*, 1270–1276.

10. Khar, A.; Lawande, K.E.; Negi, K.S. Microsatellite marker based analysis of genetic diversity in short day tropical Indian onion and cross amplification in related *Allium* spp. *Genet. Resour. Crop Evol.* **2011**, *58*, 741–752. [CrossRef]

11. Baldwin, S.; Pither-Joyce, M.; Wright, K.; Chen, L.; McCallum, J. Development of robust genomic simple sequence repeat markers for estimation of genetic diversity within and among bulb onion (*Allium cepa* L.) populations. *Mol. Breed.* **2012**, *30*, 1401–1411. [CrossRef]

12. Buckler, E.S.; Thornsberry, J.M. Plant molecular diversity and applications to genomics. *Curr. Opin. Plant Biol.* **2002**, *5*, 107–111. [CrossRef]

13. Rafalski, A. Applications of single nucleotide polymorphisms in crop genetics. *Curr. Opin. Plant Biol.* **2002**, *5*, 94–100. [CrossRef]

14. Fang, W.P.; Meinhardt, L.W.; Mischke, S.; Bellato, C.M.; Motilal, L.; Zhang, D. Accurate determination of genetic identity for a single cacao bean, using molecular markers with a nanofluidic system, ensures cocoa authentication. *J. Agric. Food Chem.* **2014**, *62*, 481–487. [CrossRef] [PubMed]

15. Cabezas, J.A.; Ibanez, J.; Lijavetzky, D.; Vélez, D.; Bravo, G.; Rodríguez, V.; Carreño, I.; Jermakow, A.M.; Carreño, J.; Ruiz-García, L.; et al. A 48 SNP set for grapevine cultivar identification. *BMC Plant Biol.* **2011**, *11*, 153. [CrossRef] [PubMed]

16. Wu, B.; Zhong, G.; Yue, J.; Yang, R.; Li, C.; Li, Y.; Zhong, Y.; Wang, X.; Jiang, B.; Zeng, J.; et al. Identification of pummelo cultivars by using a panel of 25 selected SNPs and 12 DNA segments. *PLoS ONE* **2014**, *9*, e94506. [CrossRef] [PubMed]

17. Longhi, S.; Giongo, L.; Buti, M.; Surbanovski, N.; Viola, R.; Velasco, R.; Ward, J.A.; Sargent, D.J. Molecular genetics and genomics of the rosoideae: State of the art and future perspectives. *Hortic. Res.* **2014**, *1*, 1. [CrossRef] [PubMed]

18. Fang, W.P.; Meinhardt, L.W.; Tan, H.; Zhou, L.; Mischke, S.; Zhang, D. Varietal identification of tea (*Camellia sinensis*) using nanofluidic array of single nucleotide polymorphism (SNP) markers. *Hortic. Res.* **2014**, *1*, 14035. [CrossRef] [PubMed]

19. Kang, J.H.; Yang, H.B.; Jeong, H.S.; Choe, P.; Kwon, J.K.; Kang, B.C. Single nucleotide polymorphism marker discovery from transcriptome sequencing for marker-assisted backcrossing in capsicum. *Korean J. Hortic. Sci. Technol.* **2014**, *32*, 535–543. [CrossRef]

20. Zhang, H.N.; Yong, Z.; Wei, Y.Z.; Shen, J.Y.; Lai, B.; Huang, X.M.; Ding, F.; Su, Z.X.; Chen, H.B. Transcriptomic analysis of floral initiation in litchi (*Litchi chinensis* Sonn.) based on *de novo* RNA sequencing. *Plant Cell Rep.* **2004**, *33*, 1723–1735. [CrossRef] [PubMed]

21. Patel, R.K.; Jain, M. NGS QC Toolkit: A toolkit for quality control of next generation sequencing data. *PLoS ONE* **2012**, e30619. [CrossRef] [PubMed]

22. Wang, J.; Lin, M.; Crenshaw, A.; Hutchinson, A.; Hicks, B.; Yeager, M.; Berndt, S.; Huang, W.-Y.; Hayes, R.B.; Chanock, S.J.; et al. High-throughput single nucleotide polymorphism genotyping using nanofluidic dynamic arrays. *BMC Genom.* **2009**, *10*, 561. [CrossRef] [PubMed]

23. Fluidigm. *Fluidigm SNP Genotyping User Guide Rev H1, PN 68000098*; Fluidigm Corporation: San Francisco, CA, USA, 2011.

24. Peakall, R.; Smouse, P.E. GenAlEx 6.5: Genetic analysis in excel. Population genetic software for teaching and research—An update. *Bioinformatics* **2012**, *6*, 2537–2539. [CrossRef] [PubMed]

25. Saitou, N.; Nei, M. The neighbor-joining method: A new method for reconstructing phylogenetic trees. *Mol. Biol. Evol.* **1987**, *4*, 406–425. [PubMed]

26. Pritchard, J.K.; Stephens, M.; Donnelly, P. Inference of population structure using multilocus genotype data. *Genetics* **2000**, *155*, 945–959. [PubMed]

27. Maughan, P.J.; Smith, S.M.; Rojas-Beltran, J.A.; Elzinga, D.; Raney, J.A.; Jellen, E.N.; Bonifacio, A.; Udall, J.A.; Fairbanks, D.J. Single nucleotide polymorphism identification, characterization, and linkage mapping in quinoa. *Plant Genome* **2012**, *5*, 114–125. [CrossRef]

28. Maughan, P.J.; Smith, S.M.; Fairbanks, D.J.; Jellen, E. Development, characterization, and linkage mapping of single nucleotide polymorphisms in the Grain Amaranths (Amaranthus sp.). *Plant Genome* **2011**, *4*, 92–101. [CrossRef]

29. Wang, B.; Tan, H.W.; Fang, W.; Meinhardt, L.W.; Mischke, S.; Matsumoto, T.; Zhang, D. Developing single nucleotide polymorphism (SNP) markers from transcriptome sequences for identification of longan (*Dimocarpus longan*) germplasm. *Hortic. Res.* **2015**, *2*, 14065. [CrossRef] [PubMed]

30. Zhou, L.; Vega, F.E.; Tan, H.; Lluch, A.E.R.; Meinhardt, L.W.; Fang, W.; Mischke, S.; Irish, B.; Zhang, D. Developing single nucleotide polymorphism (SNP) markers for the identification of coffee germplasm. *Trop. Plant Biol.* **2016**, *9*, 82–95. [CrossRef]