Development of an SNP marker set for marker-assisted backcrossing using genotyping-by-sequencing in tetraploid perilla

Jae-Eun Oh1,4 · Ji-Eun Kim1 · Jangmi Kim1 · Myoung-Hee Lee2 · Keunpyo Lee3 · Tae-Ho Kim3 · Sung-Hwan Jo1 · Jeong-Hee Lee1

Received: 8 August 2022 / Accepted: 26 August 2023 / Published online: 19 September 2023
© The Author(s), under exclusive licence to Springer-Verlag GmbH Germany, part of Springer Nature 2023

Abstract
High-quality molecular markers are essential for marker-assisted selection to accelerate breeding progress. Compared with diploid species, recently diverged polyploid crop species tend to have highly similar homeologous subgenomes, which is expected to limit the development of broadly applicable locus-specific single-nucleotide polymorphism (SNP) assays. Furthermore, it is particularly challenging to make genome-wide marker sets for species that lack a reference genome. Here, we report the development of a genome-wide set of kompetitive allele specific PCR (KASP) markers for marker-assisted recurrent selection (MARS) in the tetraploid minor crop perilla. To find locus-specific SNP markers across the perilla genome, we used genotyping-by-sequencing (GBS) to construct linkage maps of two F2 populations. The two resulting high-resolution linkage maps comprised 2326 and 2454 SNP markers that spanned a total genetic distance of 2133 cM across 16 linkage groups and 2169 cM across 21 linkage groups, respectively. We then obtained a final genetic map consisting of 22 linkage groups with 1123 common markers from the two genetic maps. We selected 96 genome-wide markers for MARS and confirmed the accuracy of markers in the two F2 populations using a high-throughput Fluidigm system. We confirmed that 91.8% of the SNP genotyping results from the Fluidigm assay were the same as the results obtained through GBS. These results provide a foundation for marker-assisted backcrossing and the development of new varieties of perilla.

Keywords Perilla · Allotetraploid · Genotyping-by-sequencing · Marker-assisted backcrossing · Crop breeding

Introduction
Molecular markers enable efficient crop breeding through a variety of marker-assisted selection (MAS) approaches, including marker-assisted backcrossing (MABC), marker-assisted gene pyramiding, and marker-assisted recurrent selection (MARS) (Rossetto and Henry 2014). Linkage mapping studies of family-based populations, genome-wide association studies using natural diversity panels, and joint linkage-association mapping of both biparental and natural populations are the most widely used strategies to identify the underlying genetics of crop species. Genotyping-by-sequencing (GBS) can be used to create genetic maps and identify trait-associated markers, which, once validated, can be applied in MAS to combine genomic regions of interest (Rasheed et al. 2017). Marker-assisted pyramiding of quantitative trait loci (QTLs) that individually explain a small portion of the variance in a complex trait has been largely ineffective; however, phenotypic recurrent selection can be used to improve complex traits, although the long selection cycles and difficulty in identifying unique genes or alleles restrict its practicability (Ye and Smith 2008).

MARS is a strategy that allows targeted recombination of gene effects through recurrent selection. When MARS is used to introduce a desired target trait from a donor, it is important to accurately check the recovery rate of the
background genome as well as the introduction of the target to shorten the time and increase the accuracy of inbred line development. When introducing a target trait from a wild species, it is necessary to minimize the unintentional introduction of additional wild traits by ensuring that only the locus related to the target trait is affected. Furthermore, in the case of polyploidy, it is necessary to confirm the degree of recovery by the subgenome and individual chromosomes after the new genetic material is introduced. Accordingly, a molecular marker set that can check the entire genome can be used to secure an inbred line with a high recovery rate in a relatively short time.

In order to develop MARS markers, genome sequence information must be assembled in a chromosome-level genome. Linkage maps can then be generated for segregated groups using GBS based on the assembled reference genome. These linkage maps can then be applied to select markers capable of identifying the highly similar structures of homeologous subgenomes. Polyploidy generally leads to high sequence conservation between homeologous genes, increased gene and genome dosage, and large genome size, which cause different layers of complexity for gene discovery and ultimately affect the efficient use of markers during breeding (Rasheed et al. 2017). Therefore, when developing markers in polyploid crops, it is essential to consider the redundancy present in the genome.

MARS markers are available in major crops including rice (Ahmadikiah et al. 2015), pepper (Kim et al. 2017), and tomatoes (Park et al. 2018), but not in the tetraploid minor crop perilla. Furthermore, there has been no report on the development of markers for MABC in tetraploid perilla. Perilla is an attractive crop used for oil or as a leafy vegetable because of its unique aroma and composition (Kang et al. 2019). Perilla has become increasingly popular in recent years as an economical source of omega-3 fatty acids, flavor ingredients, and flavonoids containing anthocyanins (Dhyani et al. 2019; Oh et al. 2018). There is increasing demand for molecular breeding to efficiently enhance these desirable traits in perilla varieties; however, as perilla is cultivated for food only in northeastern Asia, research on its genome and genetics is limited (Asif 2011; Sa et al. 2018). In addition, cultivars of perilla are reported to be polyploid, so it is expected that there will be difficulties in genomic research and molecular marker development (Lee et al. 2018; Nitta et al. 2005). GBS is considered an effective and widely applicable method for genotyping various crops (Hussain et al. 2017; Li et al. 2017; Bielenberg et al. 2015; Verma et al. 2015; Poland et al. 2012). Recently, a large number of single-nucleotide polymorphisms (SNPs) were successfully identified by GBS of interspecies cross-breeding groups of the diploid wild species Perilla citriodora (Kang et al. 2019).

In this report, we identified genome-wide SNPs by GBS of two F2 populations of tetraploid perilla. We then constructed a universal map composed of 22 linkage groups and selected SNP markers for each individual chromosome. Next, we selected 96 markers for MARS, converted them to allelic-specific competitive allele specific PCR (KASP) markers, and verified their accuracy in the two F2 populations using the Fluidigm BioMark HD genotyping system.

Materials and methods

DNA extraction and library preparation for GBS

In order to secure sufficient candidate SNP marker loci, we selected two F2 segregating populations by crossing P. frutescens var. crispa (Chajogi #107-1 or Pureunchajogi #115-2) and P. frutescens var. frutescens (Namcheon). One population, #973,109, was obtained by crossing Chajogi #107-1×Namcheon, and the other population, #973116, was obtained by crossing Pureunchajogi #115-2×Namcheon. After self-fertilization of the F1 generation of each population, F2 lines were cultivated at the National Institute of Crop Science, RDA (Miryang, Korea) (Oh et al. 2018).

We randomly collected 96 fresh young leaves from each F2 line and extracted genomic DNA using a modified cetyltrimethylammonium bromide-based method (Allen et al. 2006). We then used the extracted genomic DNA to prepare libraries for GBS and whole-genome sequencing. The DNA was quantified using a Thermo Scientific Nanodrop 8000 spectrophotometer (Fisher Scientific; Hampton, NH, USA). To prepare the GBS libraries, we digested the DNA using the restriction enzymes PstI (CTGCAG) and MspI (CCGG) and ligated the DNA from each F2 line to a set of 96 barcoded adapters using a modified version of a previously published protocol (Elshire et al. 2011). We then pooled the libraries from each F2 line and sequenced them in 151-bp reads using TrueSeq Version 3.0 paired-end sequencing on the Illumina platform HiSeq X Ten. Each lane of sequencing contained DNA from all 96 samples of a single F2 line.

Sequencing and SNP filtration

After sequencing, we removed the reads containing only the common adapter using the Cutadap software (Martin 2011). We then de-multiplexed the raw reads using a Python script, which filtered out reads containing ambiguous bases in the barcodes and split the raw Illumina FASTQ file into 96 separate FASTQ files based on the barcode sequences (Eun et al. 2016). Next, we trimmed the de-multiplexed reads using the Solexa QA package v.1.13 (Cox et al. 2010). Poor-quality reads and reads shorter than 25 bases were discarded. We applied the Burrows–Wheeler Aligner program
(BWA; 0.6.1-r104) (Li and Durbin 2009) to align the clean reads to the draft genome sequence of Perilla frutescens (unpublished data) for SNP calling. The high mapping quality ensured reliable mapping of the reads, which is important for variant calling. Next, we extracted the mapped reads from the resulting BAM file for further analyses using SAMtools v.0.1.16 (Li et al. 2009). Using the SAM tools varFilter command, SNPs were called only for variable positions with a minimal mapping quality (i.e., -Q) of 30. The minimum and maximum read depths were set to 3 and 100, respectively. We used an in-house script that considered biallelic loci to select the significant sites among the called SNP positions (Kang et al. 2019; Kim et al. 2014, 2018). A raw SNP matrix was constructed by collecting the raw SNP positions obtained from comparison of each sample with the draft genome. Then, the blank region (non-SNP loci) was filled from the consensus sequence of the sample, and a final SNP matrix was created by filtering out mis-called SNP loci. Based on the ratio of the mapped reads to each SNP locus, we classified the SNPs into three types: Homozygous SNPs, read depth ≥ 90%; Heterozygous SNPs, 40% ≤ read depth ≤ 60%; and Other SNPs, which did not meet the criteria for Homozygous or Heterozygous. We then selected SNPs that had minor allele frequency (MAF) > 5% and missing data in less than 30% of the samples.

Linkage map construction

We converted the filtered SNPs in the F2 populations to a genotype matrix consisting of four characters [maternal type: a; paternal type: b; heterozygous type: h; unknown: n]. We then constructed a genetic linkage map of P. frutescens var. frutescens separately using Joinmap 4 (Van Ooijen 2006). Also, we filtered out genetically identical markers and individuals using the genetic mapping program. The SNP markers were grouped into linkage groups (LGs) with logarithm of the odds threshold value ≥ 4. The genetic distances of the SNP markers, which were based on the recombination rates, were then converted using Kosambi’s mapping function (Kosambi 2016). The final genetic linkage map was drawn using MapChart ver2.3 (Voorrips 2002).

SNP selection and validation by fluidigm genotyping assay

SNP markers were selected from the markers constituting the two genetic maps. To design Fluidigm SNP genotyping assays, 60–150 bp sequences flanking the selected SNPs on either side were founded in the perilla draft genome sequence and checked for specificity using BLAST. Finally, the selected SNPs and flanking sequences were uploaded to the D3 Assay Design (https://d3.fluidigm.com/) website. After confirming the results, the designed assays were ordered. One Fluidigm SNP assay contains allele-specific primer 1 (ASP1), ASP2, locus-specific primer (LSP), and specific target amplification (STA) primer.

The SNP genotyping assays were performed in series using the BioMark™ HD system (Fluidigm, San Francisco, CA, USA) at SNP genetics (Seoul, Korea) according to the manufacturer’s instructions using the following thermal cycling conditions: 95 °C for 15 s, 64 °C for 45 s, and 72 °C for 15 s, with a touchdown of – 1 °C per cycle from 64 °C to 61 °C, followed by 34 cycles of 95 °C for 15 s, 60 °C for 45 s, and 72 °C for 15 s. The genotyping results were acquired using the Fluidigm SNP Genotyping Analysis software. All base calls were manually verified, and any errors in homozygous or heterozygous clusters were curated.

Results

Identification of genome-wide SNPs using GBS

The summary statistics of the sequencing and GBS analysis are shown in Table S1. Paired-end sequencing produced approximately 105.4 Gbp (698,076,320 reads) from #973,109 and 109.8 Gbp (727,703,068 reads) from #973,116. Following removal of the barcodes and PstI and MspI overhang sequences, 684,566,618 and 715,551,854 short reads from #973,109 and #973,116, respectively, were de-multiplexed. After elimination of adapters and ambiguous nucleotides, 629,718,268 and 660,395,486 high-quality reads with Phred quality score ≥ 20 were obtained for #973,109 and #973,116, respectively. A total of 587,492,927 reads and 622,599,718 reads from #973,109 and #973,116, respectively, were mapped to the P. frutescens var. frutescens draft genome, which accounted for about 93.2% of the reads and 94.2% of the trimmed reads from #973,109 and #973,116, respectively. The average depth of the mapped reads in the GBS data sets was 108x and 103x for #973,109 and #973,116, respectively.

SNP filtering by linkage mapping

To secure biparental polymorphic SNPs in each population, we aligned the whole-genome sequencing data of Chajogi #107-1, Pureunchajogi #115-2, and Namcheon to the P. frutescens var. frutescens draft genome and identified genome-wide SNPs. In total, we identified 3,531,724 SNPs between Chajogi #107-1 and Namcheon and 3,659,790 SNPs between Pureunchajogi #115-2 and Namcheon. (Table 1). Subsequent analysis using our in-house GBS analysis pipeline considering read depth, mapping quality, and biallelic loci (Eun et al. 2016; Kim et al. 2014; Lee et al. 2018b) identified a total of 38,156 and 44,219 raw SNPs in #973,109 and #973,116, respectively (Table 1). These SNPs
were filtered to identify putative markers using the criteria of 30% missing values across the genotyped individual and MAF ≥ 0.05, which yielded a total of 5,052 and 5,462 SNPs for #973,109 and #973,116, respectively. Next, considering the biparental polymorphic SNPs in each population, we identified 2,840 markers in #973,109 and 3,147 markers in #973,116 (Table 1). After additional filtering using the mapping program with criteria such as chi-square-test and removal of duplicated markers and individuals, 2783 and 3072 markers were converted separately to genotyping matrixes for mapping of each GBS dataset.

The genetic map of #973,109 formed 16 LGs using 2326 out of the 2783 selected SNP markers (Table S2). The number of SNP markers constituting each group ranged from 25 (LG15) to 323 (LG16). The genetic distance (cM) of the LGs ranged from 47.9 cM for LG15 to 220.5 cM for LG08. The total genetic distance was 2,133.9 cM, and the average genetic distance between markers was 0.9 cM. The genetic map of #973,116 formed 21 LGs using 2,454 out of the 3,072 selected markers (Table S3). The number of markers constituting each group ranged from 30 (LG10) to 244 (LG16). The genetic distance (cM) of the LGs ranged from 40.8 cM for LG10 to 167.0 cM for LG19. The total genetic distance was 2,169.3 cM, and the average genetic distance between markers was 1.0 cM. When we integrated the two genetic maps with the commonly used markers, nine paired LGs were matched one-to-one between the two genetic maps; however, for the remaining 19 LGs, different numbers of LGs were matched between the two maps (Table 2; Fig. 1). Several of the 19 LGs may be the result of the inability to distinguish between sub-genomes. However, by comparing the two genetic maps, the sub-genomes can be more clearly distinguished.

Marker selection for MARS

To select markers for MARS, we compared the 2,454 SNPs in #973,116 to the 2,326 SNPs in #973,109 and identified the SNPs that were common to both linkage maps. A total of 1,123 SNPs were represented in both F2 populations (Table 2; Fig. 1). In addition, we identified 556 SNP markers by comparing the genetic maps of #973,116 and #973,109 with a genetic map of the recombinant inbred line developed from a cross between Daesil and Ipdeulkkae 01 (unpublished data). We selected target loci from among the candidate SNPs by choosing those with no other SNPs or indels within 300 bp of the target region. In addition, we ensured that the selected SNP markers occurred at regular intervals along the genome so that the whole region of each LG was covered. The largest and smallest LGs had nine and two SNP markers, respectively. Thus, 96 SNP markers were appropriately assigned to individual LGs corresponding to the number of chromosomes.

High-throughput genotyping using fluidigm

We checked the accuracy of the 96 selected SNP markers by genotyping the 192 samples of the F2 lines #973,116 and #973,109 using the Fluidigm BioMark HD system. The primers used for the Fluidigm SNP assay are summarized in Table 3. Ninety-one (94.8%) of the 96 markers could successfully differentiate the ancestral alleles in

---

Table 1 Single-nucleotide polymorphism (SNP) selection statistics

| Filtration criteria                          | Number of SNPs |
|---------------------------------------------|----------------|
| #973,109                                    | #973,116       |
| No. of polymorphic SNPs between parents     | 3,531,724      | 3,659,790      |
| (from resequencing)                         |                |                |
| No. of total SNPs from F2 populations by    | 38,156         | 44,219         |
| GBS                                         |                |                |
| MAF (minor allele frequency) > 5%           | 12,744         | 15,086         |
| Missing data < 30%                          | 5052           | 5462           |
| Polymorphic SNPs between parents            | 2840           | 3147           |
| No. of SNP for genotyping                   | 2783           | 3072           |

Table 2 Statistics of single-nucleotide polymorphism (SNP) selection for MARS

| LG02 | LG07 | 74 | 7  |
| LG01 | LG05 | 12 | 3  |
| LG01 | LG20 | 54 | 3  |
| LG14 | LG21 | 27 | 5  |
| LG08 | LG01 | 100| 3  |
| LG08 | LG02 | 65 | 4  |
| LG03 | LG15 | 45 | 5  |
| LG13 | LG12 | 65 | 5  |
| LG12 | LG18 | 49 | 6  |
| LG12 | LG17 | 48 | 4  |
| LG06 | LG13 | 49 | 4  |
| LG05 | LG14 | 57 | 4  |
| LG16 | LG16 | 143| 5  |
| LG16 | LG11 | 78 | 5  |
| LG03 | LG09 | 42 | 3  |
| LG04 | LG10 | 19 | 4  |
| LG09 | LG06 | 27 | 4  |
| LG11 | LG06 | 29 | 2  |
| LG10 | LG06 | 57 | 5  |
| LG07 | LG19 | 34 | 9  |
| LG05 | LG04 | 31 | 4  |
| LG05 | LG08 | 18 | 2  |
| Sum  |      | 1123| 96 |

---
the F2 generation, although three of those had a call rate of 100 but a confidence value of 0 (Fig. 2). Out of a total of 18,432 high-throughput genotyping reactions, 17,403 reactions were successful, whereas 1029 reactions were unsuccessful. In order to confirm whether the Fluidigm SNP genotyping results were the same as the GBS results, we identified 15,935 experiments present in both sets of genotyping results (Table 4). Of these common experiments, 14,658 (91.8%) gave the same result with both genotyping methods, whereas 1277 (9.1%) gave different results. When we investigated the results in both F2 lines, 191 reactions (15.2%) were Heterozygous-type markers in the GBS results but Homozygous-type markers in the Fluidigm results. Conversely, 1071 reactions (83.9%) were Homozygous-type markers in the GBS results but heterozygous-type markers in the Fluidigm results (Table 4).

Discussion

MAS makes crop breeding programs faster and more accurate, especially if the life cycle of the crop is long, or there are not enough research personnel for the target crop. When multiple genes are introduced into different chromosomes, difficulties may arise due to unintentional dragging of nearby genes. Therefore, markers are needed that cover all the chromosomes evenly in order to investigate the recovery of the genetic background after new genes are introduced. This can be done either with a chromosome-level reference genome or with a complete linkage map of the genome. Although new reference genomes are being reported at a rapid rate, there are still many crops, especially minor crops, for which chromosome-level reference genomes are not available. It is always desirable to have a reference genome when designing MARS markers; however, it can be time consuming and costly to create a new high-quality reference genome. As an alternative approach, complete linkage maps can be used to develop MARS marker sets for crops lacking a suitable reference genome.

To create a set of MARS markers for perilla, which is a ploidy orphan crop with sparse background research and no available chromosome-level reference genome, we made a linkage map using F2 populations. It is particularly challenging to select allelic-specific SNP markers that work well in polyploid crops, because it is not possible to determine how much similarity exists between subgenomes, especially in the absence of a high-quality reference genome. An important and difficult attempt to effectively distinguish each subgenome in polyploids such as bread wheat was made previously (Makhoul et al. 2020). Even with these difficulties, we were able to select appropriate candidate markers through creation of a linkage map. Although the 20 LGs, which equaled the number of perilla chromosomes, were not fully identified, a high success rate was confirmed in the experimental verification process while covering the entire chromosome.

The number of LGs differed depending on the population used to create the genetic map. Sixteen LGs were formed in #973,109, whereas 21 LGs were formed in #973,116 (Table S1 and Table S2). When the corresponding LGs were compared on the genetic maps of the two populations, LG01 in #973,109 was divided into LG05 and LG20 in #973,116. Conversely, LG06 in #973,116 was divided into LG09, LG10, and LG11 in #973,109 (Table 3). Combining two LGs into one LG or dividing one LG into two LGs is expected to fail to accurately separate the LG because of chromosomes.
Table 3  Primer sequences used for 96 SNP assay

| SNP_NAME  | ASP1_SEQ1) | ASP2_SEQ2) | LSP_SEQ3) | STA_SEQ4) | Chromosome number5) | Allele |
|-----------|------------|------------|-----------|-----------|---------------------|--------|
| PfSNP01   | GGTGTTCTGTAGTGTTCTTGTCTT | GGTGTTCTGTAGTGTTCTTGTCTT | TGGTAAAAACATGGGGCTGCTT | GGCGATAACAGGCTGCAAT | Chr01   | A/G  |
| PfSNP02   | CTTTTTGGATAAATGACATCATCTAGTTTCTT | CTTTTTGGATAAATGACATCATCTAGTTTCTT | CACCCCTGCACTTCAGCATCAAA | TCCTGGTATTCTTGGTGAACCT | Chr01   | C/A  |
| PfSNP03   | CTACTCTGATATTACCTCAGGTATATGATCATA | CTACTCTGATATTACCTCAGGTATATGATCATA | ACAACGACACACCCAGGAAGGA | CCTACGGAAGAGATCCAAATG | Chr01   | A/G  |
| PfSNP04   | TGTAGTGGCTTGGTCTTACCTGACA | TGTAGTGGCTTGGTCTTACCTGACA | CATGACAGGCCACTGACACTGA | TCGATGGTGGATTTGGTCGCC | Chr01   | A/G  |
| PfSNP05   | GCCTGAGAGATCAGAAGAGATCATGCA | GCCTGAGAGATCAGAAGAGATCATGCA | GACCTTTGCAAAACCCGATAT | ACACGCTGTGACAAATATAT | Chr01   | A/C  |
| PfSNP06   | TCTCTTCCATCAAGAGTTGGATCCTAC | TCTCTTCCATCAAGAGTTGGATCCTAC | GATCCATCAAGAGAGATGGTT | GATCCATCAAGAGAGATGGTT | Chr02   | A/G  |
| PfSNP07   | AGCAATTCTCTTATTTGGTAAAGACCTG | AGCAATTCTCTTATTTGGTAAAGACCTG | ACATAAAGTTGAAAAATTGGA | TGAACACTTGGA | Chr02   | C/A  |
| PfSNP08   | CGGCAGACATTGGGAGTATAA | CGGCAGACATTGGGAGTATAA | GGAAGCTGGCTGCACTTCACCA | TCCTGGGATATTGGTGAGC | Chr02   | T/C  |
| PfSNP09   | CACCTCACTGCCACCCATCCAT | CACCTCACTGCCACCCATCCAT | TCGGGCTGGTTGACTTCTTAT | CATGCTCAGACACAGG | Chr02   | C/A  |
| PfSNP10   | GGATCCAGTCTGCTTCCATGCTC | GGATCCAGTCTGCTTCCATGCTC | TCTCCTGTGGAGTCCATGATG | TCTCCAACTCTCACTGAGGAT | Chr02   | A/T  |
| PfSNP11   | CGGAATCCGCCAGCCTGCTA | CGGAATCCGCCAGCCTGCTA | CAGAAGCCCCCGCTTCTCCATGCTA | ATGACCTGATATACCATCAG | Chr20   | A/C  |
| PfSNP12   | TTAACAGTATGATCACAACCACCTACATCATTACACCT | TTAACAGTATGATCACAACCACCTACATCATTACACCT | ACAAACATGATCAAGATCAGTCTC | TCAGATTTTCTACTATTCCGATGT | Chr20   | A/C  |
| PfSNP13   | TTCTTGGAGATGAAAGACCGAGCA | TTCTTGGAGATGAAAGACCGAGCA | CCCACTCAGTGACACCAA | TGCTTTTCTGGAAGAGTGGTTT | Chr20   | T/A  |
| PfSNP14   | GCTGTGACTGCTATTTGATATCATTACCT | GCTGTGACTGCTATTTGATATCATTACCT | GCTTCAGATTATGGAACATTTT | TTCTATTGCAATATTCTCACA | Chr20   | T/C  |
| PfSNP15   | GCTTTGTGGTGAGCTCTGGGATTTG | GCTTTGTGGTGAGCTCTGGGATTTG | CAGACGCTGGAGGAAGAACAGA | GCACATGAGTGGTGACAGGGA | Chr05   | T/C  |
| PfSNP16   | CCTAAGGAGATTGATATAACCCAGAGCAGG | CCTAAGGAGATTGATATAACCCAGAGCAGG | AGACAAGGCCACACTGCTTCT | GCTTCCCAATCTTGTTAAATAT | Chr05   | C/G  |
| PfSNP17   | GGGATCCTCCCTCTTCTTTGTGTTG | GGGATCCTCCCTCTTCTTTGTGTTG | TGGTCTAGGAGAGGGGAGCTT | AGTGTTGCCCTACTTTGTGATAT | Chr05   | G/C  |
| PfSNP18   | CCTAGGAGGTCAGTAAACAAAAAATGAGAT | CCTAGGAGGTCAGTAAACAAAAAATGAGAT | CACGATCAGCATGCACAA | GTGGAAGTGTCTACTTTTAA | Chr05   | T/A  |
| PfSNP19   | CACCTCTGCTTGGGATATTCCATATTT | CACCTCTGCTTGGGATATTCCATATTT | TCTTCTGTGGTGGTGAAGGCC | TCGAATCCATAGTGGTATG | Chr05   | T/C  |
| PfSNP20   | GTCATGTCCTCTGAGATCAATGG | GTCATGTCCTCTGAGATCAATGG | TGGAAACGGCTACTGAGGACAG | GCAGGGAGAAAGTGTTGATTCAAA | Chr04   | G/C  |
| PfSNP21   | GACAAATCATGCTCTCTCAATATGAT | GACAAATCATGCTCTCTCAATATGAT | AGACTGGGAGACTTTTTCCGAG | TGTGATGCTGATATCCGGTTT | Chr04   | T/C  |
Table 3 (continued)

| SNP_NAME | ASP1_SEQ | ASP2_SEQ | LSP_SEQ | STA_SEQ | Chromosome number | Allele |
|----------|----------|----------|---------|---------|------------------|--------|
| PISNP22  | CTTATTCGGGCTCTCAGTCA | CTTATTCGGGCTCTCAGTCA | CTTACTAGGCAACTACAGACTGACC | GATTTGTTTGCGTAACATCA | Chr04 | A/T |
| PISNP23  | CCAAATGCTGATATCTCTCTCTCTCATGG | CTTATTCGGGCTCTCAGTCA | CTTACTAGGCAACTACAGACTGACC | GATTTGTTTGCGTAACATCA | Chr04 | G/T |
| PISNP24  | GCAATGTGGATATCTCTCTCTCTCATGG | CTTATTCGGGCTCTCAGTCA | CTTACTAGGCAACTACAGACTGACC | GATTTGTTTGCGTAACATCA | Chr04 | A/T |
| PISNP25  | CCAAATGCTGATATCTCTCTCTCTCATGG | CTTATTCGGGCTCTCAGTCA | CTTACTAGGCAACTACAGACTGACC | GATTTGTTTGCGTAACATCA | Chr04 | C/T |
| PISNP26  | GCAATGTGGATATCTCTCTCTCTCATGG | CTTATTCGGGCTCTCAGTCA | CTTACTAGGCAACTACAGACTGACC | GATTTGTTTGCGTAACATCA | Chr04 | C/A |
| PISNP27  | GCAATGTGGATATCTCTCTCTCTCATGG | CTTATTCGGGCTCTCAGTCA | CTTACTAGGCAACTACAGACTGACC | GATTTGTTTGCGTAACATCA | Chr04 | A/G |
| PISNP28  | CCAAATGCTGATATCTCTCTCTCTCATGG | CTTATTCGGGCTCTCAGTCA | CTTACTAGGCAACTACAGACTGACC | GATTTGTTTGCGTAACATCA | Chr04 | G/T |
| PISNP29  | GCAATGTGGATATCTCTCTCTCTCATGG | CTTATTCGGGCTCTCAGTCA | CTTACTAGGCAACTACAGACTGACC | GATTTGTTTGCGTAACATCA | Chr04 | A/G |
| PISNP30  | GCAATGTGGATATCTCTCTCTCTCATGG | CTTATTCGGGCTCTCAGTCA | CTTACTAGGCAACTACAGACTGACC | GATTTGTTTGCGTAACATCA | Chr04 | G/C |
| PISNP31  | GCAATGTGGATATCTCTCTCTCTCATGG | CTTATTCGGGCTCTCAGTCA | CTTACTAGGCAACTACAGACTGACC | GATTTGTTTGCGTAACATCA | Chr04 | C/T |
| PISNP32  | GCAATGTGGATATCTCTCTCTCTCATGG | CTTATTCGGGCTCTCAGTCA | CTTACTAGGCAACTACAGACTGACC | GATTTGTTTGCGTAACATCA | Chr04 | G/A |
| PISNP33  | GCAATGTGGATATCTCTCTCTCTCATGG | CTTATTCGGGCTCTCAGTCA | CTTACTAGGCAACTACAGACTGACC | GATTTGTTTGCGTAACATCA | Chr04 | T/A |
| PISNP34  | GCAATGTGGATATCTCTCTCTCTCATGG | CTTATTCGGGCTCTCAGTCA | CTTACTAGGCAACTACAGACTGACC | GATTTGTTTGCGTAACATCA | Chr04 | A/A |
| PISNP35  | GCAATGTGGATATCTCTCTCTCTCATGG | CTTATTCGGGCTCTCAGTCA | CTTACTAGGCAACTACAGACTGACC | GATTTGTTTGCGTAACATCA | Chr04 | C/T |
| PISNP36  | GCAATGTGGATATCTCTCTCTCTCATGG | CTTATTCGGGCTCTCAGTCA | CTTACTAGGCAACTACAGACTGACC | GATTTGTTTGCGTAACATCA | Chr04 | G/A |
| PISNP37  | GCAATGTGGATATCTCTCTCTCTCATGG | CTTATTCGGGCTCTCAGTCA | CTTACTAGGCAACTACAGACTGACC | GATTTGTTTGCGTAACATCA | Chr04 | A/G |
| PISNP38  | GCAATGTGGATATCTCTCTCTCTCATGG | CTTATTCGGGCTCTCAGTCA | CTTACTAGGCAACTACAGACTGACC | GATTTGTTTGCGTAACATCA | Chr04 | G/A |
| PISNP39  | GCAATGTGGATATCTCTCTCTCTCATGG | CTTATTCGGGCTCTCAGTCA | CTTACTAGGCAACTACAGACTGACC | GATTTGTTTGCGTAACATCA | Chr04 | A/G |
| PISNP40  | GCAATGTGGATATCTCTCTCTCTCATGG | CTTATTCGGGCTCTCAGTCA | CTTACTAGGCAACTACAGACTGACC | GATTTGTTTGCGTAACATCA | Chr04 | C/G |
| PISNP41  | GCAATGTGGATATCTCTCTCTCTCATGG | CTTATTCGGGCTCTCAGTCA | CTTACTAGGCAACTACAGACTGACC | GATTTGTTTGCGTAACATCA | Chr04 | T/C |
| PISNP42  | GCAATGTGGATATCTCTCTCTCTCATGG | CTTATTCGGGCTCTCAGTCA | CTTACTAGGCAACTACAGACTGACC | GATTTGTTTGCGTAACATCA | Chr04 | A/G |
| SNP_NAME   | ASP1_SEQ (1) | ASP2_SEQ (2) | LSP_SEQ (3) | STA_SEQ (4) | Chrosome number (5) | Allele |
|------------|--------------|--------------|-------------|-------------|---------------------|--------|
| PfSNP43    | TGAACGCCGCATTCCA | GAACGCAGGCATTCCG | AGTTAATTGTATTTAAAGAAATAAA | GITCGGAAATGAAATAAGTTTG AAACCC | Chr08   | T/C     |
| PfSNP44    | TGACGAGGGAAACCTCGC | ATGACGGAGGGAAACCTCGC | TTATGTAATGACCTACATAT | GTATTTGGCATATCTTTGGAGG | Chr16   | C/T     |
| PfSNP45    | GTGAAATGTTGACAGATTGGA | TGAATATGTTGACAGATTGGA | TACACAAATGGCTCGTTGGA | TAGGATGCGGTAGGTT | Chr16   | T/C     |
| PfSNP46    | CAATATTAATATCATTCGAGC | CAATATTAATATCATTCGAGC | AGTCTGATGCTTTGTAAGGAA GCA | CTGTTGCGAACGGACC | Chr16   | A/G     |
| PfSNP47    | ACTCCAGAGATTCAAATCTCAGTAC | ACTCCAGAGATTCAAATCTCAGTAC | CTGCAATGCGACCATTACGCT | GCACGATTCGAAATACACTTG AAAT | Chr16   | C/T     |
| PfSNP48    | GCCGAGAATCTTCCTCCAGA AAGAT | GCCGAGAATCTTCCTCCAGA AAGAT | AGT TTGAAAGGAAG | TTGCAACATGCCTCCTACAG | Chr10   | T/C     |
| PfSNP49    | CCATGCTTTCCTCATATCTTCTT ATCCAGA | CCATGCTTTCCTCATATCTTCTT ATCCAGA | AGT TTGAAAGGAAG | TTGCAACATGCCTCCTACAG | Chr10   | T/C     |
| PfSNP50    | ACCGAAATGAGAAACAGATGGGA | ACCGAAATGAGAAACAGATGGGA | AGT TTGAAAGGAAG | TTGCAACATGCCTCCTACAG | Chr10   | T/G     |
| PfSNP51    | ATCGGGCTGCTAAAGACGGTGA | ATCGGGCTGCTAAAGACGGTGA | AGT TTGAAAGGAAG | TTGCAACATGCCTCCTACAG | Chr10   | A/G     |
| PfSNP52    | CCCGCAATTCTCCCTCGGAA | CCCGCAATTCTCCCTCGGAA | AGT TTGAAAGGAAG | TTGCAACATGCCTCCTACAG | Chr10   | A/G     |
| PfSNP53    | GGTTCGTTTTGTGATGAAAGGA | GGTTCGTTTTGTGATGAAAGGA | AGT TTGAAAGGAAG | TTGCAACATGCCTCCTACAG | Chr10   | T/C     |
| PfSNP54    | AGAAGCTCCAGTGTTGATCA AGATA | AGAAGCTCCAGTGTTGATCA AGATA | AGT TTGAAAGGAAG | TTGCAACATGCCTCCTACAG | Chr13   | T/G     |
| PfSNP55    | GAATTGATAGTGATTTTCCTGC AGCTC | GAATTGATAGTGATTTTCCTGC AGCTC | AGT TTGAAAGGAAG | TTGCAACATGCCTCCTACAG | Chr13   | C/T     |
| PfSNP56    | GACCTTTAAATGCGAAAAAGAGA AAGGA | GACCTTTAAATGCGAAAAAGAGA AAGGA | AGT TTGAAAGGAAG | TTGCAACATGCCTCCTACAG | Chr13   | G/C     |
| PfSNP57    | GCCGGGTGTTGTTCTCTACATC | GCCGGGTGTTGTTCTCTACATC | AGT TTGAAAGGAAG | TTGCAACATGCCTCCTACAG | Chr13   | T/C     |
| PfSNP58    | CGGCAACTCTCTCTCTCTCCAGTA | CGGCAACTCTCTCTCTCTCCAGTA | AGT TTGAAAGGAAG | TTGCAACATGCCTCCTACAG | Chr07   | T/C     |
| PfSNP59    | GATTCATCCAGAAGAAACATA TCCCTACCTTAC | GATTCATCCAGAAGAAACATA TCCCTACCTTAC | AGT TTGAAAGGAAG | TTGCAACATGCCTCCTACAG | Chr07   | C/G     |
| PfSNP60    | GCCACCGACGCGG | GCCACCGACGCGG | AGT TTGAAAGGAAG | TTGCAACATGCCTCCTACAG | Chr07   | G/A     |
| PfSNP61    | GCCGATTCTACTTTCCAAAGGG | GCCGATTCTACTTTCCAAAGGG | AGAAGACCGTGACCACCTGCTT | GCCGATTCTACTTTCCAAAGGG | Chr19   | G/C     |
| PfSNP62    | GAGTCTCGTTGTTGATTCAGT | GAGTCTCGTTGTTGATTCAGT | AGT TTGAAAGGAAG | TTGCAACATGCCTCCTACAG | Chr19   | A/G     |
| PfSNP63    | CCAGACATCATCTGCATTTTG TAAACA | CCAGACATCATCTGCATTTTG TAAACA | AGT TTGAAAGGAAG | TTGCAACATGCCTCCTACAG | Chr19   | A/T     |
| SNP_NAME | ASP1_SEQ | ASP2_SEQ | LSP_SEQ | STA_SEQ | Chromosome number | Allele |
|----------|----------|----------|---------|---------|------------------|--------|
| PISNP64  | TGGAACTGTCAAGTTTCTAG   | TGGAACTGTCAAGTTTCTAG   | ACCTCTATCAGTCAGGGG  | TGTCAAATGCAACCCTCAGC  | Chr19   | C/T    |
| PISNP65  | GCTGCGCTCTATCAAC    | TGCTGCAGCCTCATCAAAT   | CTCTTTTTTTTTCT    | CTCTTTTTTTTTCT    | Chr12   | C/T    |
| PISNP66  | GAGCGTACACCTACAAGCCT | GACCGTACACCTACAAGCCT | CACCGCGTACACCTACAAGCCT | CACCGCGTACACCTACAAGCCT | Chr12   | T/C    |
| PISNP67  | CAGTTTATCGACCTGACAT | CAGTTTATCGACCTGACAT | ATTTTTTTTTTTTTTTT | ATTTTTTTTTTTTTTTT | Chr12   | C/A    |
| PISNP68  | TTTCTTGGAGGAAGGATACG | TTTCTTGGAGGAAGGATACG | ATTTTTTTTTTTTTTTT | ATTTTTTTTTTTTTTTT | Chr12   | A/G    |
| PISNP69  | TGATGTTACTTTAGTTGATT | TGATGTTACTTTAGTTGATT | ATTTTTTTTTTTTTTTT | ATTTTTTTTTTTTTTTT | Chr12   | T/C    |
| PISNP70  | CACTCAAGGTACTTCTTAAAT | CACTCAAGGTACTTCTTAAAT | ATTTTTTTTTTTTTTTT | ATTTTTTTTTTTTTTTT | Chr12   | A/C    |
| PISNP71  | CACCAACCGTAAACGCCGATA | CACCAACCGTAAACGCCGATA | ATTTTTTTTTTTTTTTT | ATTTTTTTTTTTTTTTT | Chr11   | A/G    |
| PISNP72  | CTGGGTTACCCCTACAGCAAC | CTGGGTTACCCCTACAGCAAC | ATTTTTTTTTTTTTTTT | ATTTTTTTTTTTTTTTT | Chr11   | A/G    |
| PISNP73  | ACTAAAGCTGACATTGGTCT | ACTAAAGCTGACATTGGTCT | ATTTTTTTTTTTTTTTT | ATTTTTTTTTTTTTTTT | Chr11   | A/G    |
| PISNP74  | CACCAACACAGAATCTACTG | CACCAACACAGAATCTACTG | ATTTTTTTTTTTTTTTT | ATTTTTTTTTTTTTTTT | Chr11   | A/G    |
| PISNP75  | AGGAAATGCAAATCAGTGGAA | AGGAAATGCAAATCAGTGGAA | ATTTTTTTTTTTTTTTT | ATTTTTTTTTTTTTTTT | Chr11   | C/T    |
| PISNP76  | CCAATTTATATGCTCTACCTG | CCAATTTATATGCTCTACCTG | ATTTTTTTTTTTTTTTT | ATTTTTTTTTTTTTTTT | Chr11   | C/T    |
| PISNP77  | GATCAACATTATGGTTGGG | GATCAACATTATGGTTGGG | ATTTTTTTTTTTTTTTT | ATTTTTTTTTTTTTTTT | Chr11   | A/G    |
| PISNP78  | GATGGAAACACCGGGGCTTA | GATGGAAACACCGGGGCTTA | ATTTTTTTTTTTTTTTT | ATTTTTTTTTTTTTTTT | Chr11   | T/C    |
| PISNP79  | GTCCGTTGAATGATGTGCTCCT | GTCCGTTGAATGATGTGCTCCT | ATTTTTTTTTTTTTTTT | ATTTTTTTTTTTTTTTT | Chr11   | T/A    |
| PISNP80  | AGACATTGTTAGCTCATCACAAT | AGACATTGTTAGCTCATCACAAT | ATTTTTTTTTTTTTTTT | ATTTTTTTTTTTTTTTT | Chr11   | A/C    |
| PISNP81  | CTCCAGGTGATATAAAGAGGAG | CTCCAGGTGATATAAAGAGGAG | ATTTTTTTTTTTTTTTT | ATTTTTTTTTTTTTTTT | Chr14   | C/T    |
| PISNP82  | GCACAAAGGGCAAGTAAATTAAC | GCACAAAGGGCAAGTAAATTAAC | ATTTTTTTTTTTTTTTT | ATTTTTTTTTTTTTTTT | Chr14   | C/T    |
| PISNP83  | TGCAAAAGGATGATGATGATG | TGCAAAAGGATGATGATGATG | ATTTTTTTTTTTTTTTT | ATTTTTTTTTTTTTTTT | Chr14   | T/C    |
| SNP_NAME  | ASP1_SEQ | ASP2_SEQ | LSP_SEQ | STA_SEQ | Chromosome number | Allele |
|-----------|----------|----------|---------|---------|-------------------|--------|
| PfSNP84   | AAGACCGATAAGGACGTG AACA | AAGACGGATAAGGACGTG AACA | ACTCACTGAGCTTGCAGC | AATCATGAAAGAGACGATTTT TCGGA | Chr14   | T/C       |
| PfSNP85   | CGACTCTAGAGACTGCGAC CACCTG | CGACTCTAGAGACTGCGAC CACCTG | ACGACGGCAAGAGACCTCCA | TGCAAGGCTGCCACCG | Chr15   | C/G       |
| PfSNP86   | GCTTTGGTTTTGTGGATGAT CACAA | GCTTTGGTTTTGTGGATGAT CACAA | GGCCTTTGAACCTGAGCAGC | CCCTCTTTGATCTGATGTC | Chr15   | T/C       |
| PfSNP87   | CAAGATGAAAGCGGACATCATCT GTA | CAAGATGAAAGCGGACATCATCT GTA | ACGAAGGACAGACATCGGAA | TCTTTCTGAAAGATTTTTAC GCCC | Chr15   | T/A       |
| PfSNP88   | CCGCTACCTAGCAAAAAATTCT ACGAGC | CCGCTACCTAGCAAAAAATTCT ACGAGC | GCTTTAATGTATCTACTT | CCCAAAATCGACACCGTC | Chr15   | A/T       |
| PfSNP89   | GACGTGATTTTCGGGTACAAA ATGAT | GACGTGATTTTCGGGTACAAA ATGAT | TGGTTACGTATATCTACTTT | GCAGTTCCATGCGGAAATTG | Chr17   | A/G       |
| PfSNP90   | GTAACAAAAAGAAAAAGAAAG AATAAGAACAGG | GTAACAAAAAGAAAAAGAAAG AATAAGAACAGG | TCTAGAATTATATAATCTCCA | AGCTGGCCAAGAGAAA | Chr17   | G/T       |
| PfSNP91   | CAGCTTCAGTTGACCGGCT | CAGCTTCAGTTGACCGGCT | CATGACAGATCTGACAGAAT CCA | CATGACAGATCTGACAGAAT CCA | Chr01   | T/C       |
| PfSNP92   | CGGCTTAATAACATCCCATGT AACTC | CGGCTTAATAACATCCCATGT AACTC | GCTTTCTGAGGCAACAT | CTGACATCGTACCGCCCT | Chr01   | G/A       |
| PfSNP93   | CTTAAGGACGCCGACTTTTTATGTA | CTTAAGGACGCCGACTTTTTATGTA | GCCAACAGCTTGACCCAT | GCCGTTAAACGATAAGCTACTGC | Chr02   | A/C       |
| PfSNP94   | AGAAGAATGACTAGACATTCAA ATACCCGCG | AGAAGAATGACTAGACATTCAA ATACCCGCG | GCCACACCTTGGATATTTTG | TCTTACGTCGTTGTATCTGCA | Chr20   | C/A       |
| PfSNP95   | GGGACACAGGACACCACACGC CACA | GGGACACAGGACACCACACGC CACA | GCCACACCTGGAGAAATTTTG | AGGACACCTTGGAGACACACGC | Chr05   | T/G       |
| PfSNP96   | GGGACACACAGGACACCACACGC CACA | GGGACACACAGGACACCACACGC CACA | GCCACACCTGGAGAAATTTTG | AGGACACCTTGGAGACACACGC | Chr05   | T/G       |

1) ASP1_SEQ: Allele-specific primer 1 sequence  
2) ASP2_SEQ: Allele-specific primer 2 sequence  
3) LSP_SEQ: Locus-specific primer sequence  
4) STA_SEQ: Specific target amplification primer sequence  
5) Chromosome number: Cited Tamura K et al., (2023)
with high similarity between subgenomes. This problem can be solved through genome assembly using long-read sequencing and comparison with genetic maps. The chromosome-level genome sequence was recently reported (Tamura et al. 2023). Thus the 22 linkage groups were compared to the recently published chromosome-level genome sequence. As a result of the comparison, in most case, one chromosome was matched to one linkage group. However each of the three chromosomes corresponded by dividing into two linkage groups respectively. That is chromosome 2 corresponded to LG02 and LG03, chromosome 5 to LG05 and LG06, and chromosome 12 to LG17 and LG18. On the other hand, there were cases where two chromosomes 14 and 18 corresponded to one linkage group, LG20 (Table S4). The genome sequence and the markers in the linkage group showed high synteny.

The application of PCR-based genotyping is essential for the rapid utilization of molecular markers in practical MAS breeding programs. We confirmed that the marker conversion from GBS markers to PCR-based KASP markers had a high success rate of over 91%. Thus, by applying the genetic map to perilla without a chromosome-level reference genome, we secured a molecular marker set that covered the entire genome. Our results show how a genetic map can be used to classify subgenomes in tetraploid crops. When we confirmed the converted SNP markers in the F2 lines, 8.1% of the genotyping results for the markers were different between the GBS and the Fluidigm assay. If we look at the two sets of genotyping results in more detail, we find that certain markers were more likely than others to differ between the two methods. Hence, this type of comparison may be a good way to check the accuracy of markers in the marker selection process.

**Supplementary Information** The online version contains supplementary material available at https://doi.org/10.1007/s00438-023-02066-6.

**Author contribution** JO, JE, SH, and JH conceived and designed the experiments; JO and JE conducted the SNP analysis, linkage map

---

**Fig. 2** SNP genotyping of 96 MARC markers with Fluidigm. Bars represent the SNP genotype ratios in the #973109 (a) and #973116 (b) F2 populations. The solid line shows the auto confidence of 96 markers

**Table 4** SNP genotyping comparison of 96 markers using Fluidigm and GBS

| Sum. of SNP assay | 18,432 | - | - |
|------------------|--------|---|---|
| No. of rxn for No Call in Fluidigm or GBS | 2,497 | - | - |
| No. of rxn to comparable Fluidigm and GBS results | 15,935 | 100 % | - |
| Identical genotyping in Fluidigm & GBS | 14,658 | 91.9 % | - |
| Different genotyping in Fluidigm & GBS | 1,277 | 9.1 % | 100 % |
| - Case1: No. of rxn where Fluidigm (He) and GBS (Ho) | 1,071 | - | 83.9 % |
| - Case2: No. of rxn where Fluidigm (Ho) and GBS (He) | 191 | - | 15.2 % |
| - Case3: No. of rxn where Fluidigm (Ho) and GBS (Ho) | 12 | - | 0.9 % |
construction, and data analysis; JK constructed the GBS library; MH developed the mapping population; TH provided the draft genome; KL participated in discussions about the experiments; SH and JH wrote and reviewed the manuscript.

**Funding** This work was supported by a grant from the National Agricultural Genome Project (No. PJ013355032021), Rural Development Administration, Republic of Korea.

**Data availability** The datasets generated during and/or analysed during the current study are available in the NABIC (National Agricultural Biotechnology Information Center, https://nabic.rda.go.kr/) repository, [NN-5662 ~ 5712].

**References**

Ahmadikahh A, Mirarab M, Palevani MH, Navyeripasand L (2015) Marker-assisted backcrossing to develop an elite cytoplasmic male sterility line in rice. Plant Genome. https://doi.org/10.3835/plant genome2014.07.0031

Ahmed HM (2018) Ethnomedicinal, phytochemical and pharmacological investigations of *Perilla frutescens* (L.) britt. Molecules. https://doi.org/10.3390/molecules24010102

Allen GC, Flores-Vergara MA, Krasyanski S, Kumar S, Thompson WF (2006) A modified protocol for rapid dna isolation from plant tissues using cetyltrimethylammonium bromide. Nat Protoc 1(5):2320–2325. https://doi.org/10.1038/nprot.2006.384

Asif M (2011) Health effects of omega-3,6,9 fatty acids: *Perilla frutescens* is a good example of plant oils. Orient Pharm Exp Med 11(1):51–59. https://doi.org/10.3396/13596-011-0002-x

Bielenberg DG, Rauh B, Fan S, Gasic K, Abbott AG, Reighard GL, Okie WR, Wells CE (2015) Genotyping by sequencing for SNP-based linkage map construction and QTL analysis of chilling requirement and bloom date in peach [Prunus persica (L.) batsch]. PLoS One 10(10):e0139406

Cox MP, Peterson DA, Biggs PJ (2010) SolexaQA: At-a-glance quality assessment of illumina second-generation sequencing data. BMC Bioinformatics 11:485. https://doi.org/10.1186/1471-2105-11-485

Dhyani A, Chopra R, Garg M (2019) A review on nutritional value, functional properties and pharmaceutical application of perilla (*Perilla frutescens* L.), Biomed Pharmacol J 12(2):649–660. https://doi.org/10.13005/bpj/1685

Elshir RJ, Glauhtic JC, Sun Q, Poland JA, Kawamoto K, Buckler ES, Mitchell SE (2011) A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. PLoS One 6(5):e19379. https://doi.org/10.1371/journal.pone.0019379

Eun MH, Han JH, Yoon JB, Lee J (2016) QTL mapping of resistance to the cucumber mosaic virus P1 strain in pepper using a genotyping-by-sequencing approach. Hortic Environ Biotechnol 57(6):589–597. https://doi.org/10.1007/s13580-016-0128-3

Hussain W, Baenzigser PS, Belakmar V, Gutierrez MJ, Venegas JP, Eastlery A, Sallam A, Poland J (2017) Genotyping-by-sequencing derived high-density linkage map and its application to qtl mapping of flag leaf traits in bread wheat. Sci Rep 7(1):16394. https://doi.org/10.1038/s41598-017-16006-z

Kang YJ, Lee BM, Nam M, Oh KW, Lee MH, Kim TH, Jo SH, Lee JH (2019) Identification of quantitative trait loci associated with flowering time in perilla using genotyping-by-sequencing. Mol Biol Rep 46(4):4397–4407. https://doi.org/10.1007/s11033-019-04894-5

Kim JE, Oh SK, Lee JH, Lee BM, Jo SH (2014) Genome-wide SNP calling using next generation sequencing data in tomato. Mol Cells 37(1):36–42. https://doi.org/10.14348/molecules.2014.2241

Kim H, Yoon JB, Lee J (2017) Development of fluidigm SNP type genotyping assays for marker-assisted breeding of chili pepper (*Capsicum annuum* L.). Korean J Hortic Sci Technol 35(4):465–479

Kim WJ, Ryu J, Im J, Kim SH, Kang SY, Lee JH, Jo SH, Ha BK (2018) Molecular characterization of proton beam-induced mutations in soybean using genotyping-by-sequencing. Mol Genet Genomics 293(5):1169–1180. https://doi.org/10.1007/s00438-018-1448-z

Kosambi DD (2016) The estimation of map distances from recombination values. Dd kosambi. Springer. p. 125–130. https://doi.org/10.1007/1469-1809.1943.tb02321.x

Lee MH, Oh KW, Kim MS, Kim SU, Kim JH, Oh EY, Pae SB, Yeo US, Kim T-H, Lee JH (2018) Detection of QTLs in an interspecific cross between *Perilla citriodora* P. hirtella mapping population. Korean J Breed Sci 50(1):13–20

Li H, Durbin R (2009) Fast and accurate short read alignment with burrows-wheeler transform. Bioinformatics 25(14):1754–1760. https://doi.org/10.1093/bioinformatics/btp324

Li L, Zhao S, Su J, Fan S, Pang C, Wei H, Wang H, Gu L, Zhang C, Liu G et al (2017) High-density genetic linkage map construction by F2 populations and QTL analysis of early-maturity traits in upland cotton (*Gossypium hirsutum* L.). PLoS One 12(8):e0182918. https://doi.org/10.1371/journal.pone.0182918

Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, Genome Project Data Processing S (2009) The sequence alignment/map format and SAMtools. Bioinformatics 25(16):2078–2079. https://doi.org/10.1093/bioinformatics/btp352

Makhoul M, Rambila C, Voss-Fels K, Hickey L, Snowdon R, Obermeier C (2020) Overcoming polyploidy pitfalls: a user guide for effective SNP conversion into KASP markers in wheat. Theor Appl Genet 133:2433–2430. https://doi.org/10.1007/s00122-020-04060-x

Martin M (2011) Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet J 17(1):3. https://doi.org/10.14806/ej.17.1.200

Nitta M, Lee JK, Kang CW, Katsuta M, Yasumoto S, Liu D, Nagamine T, Ohnishi O (2005) The distribution of perilla species. Genet Resour Crop Evol 52(7):797–804. https://doi.org/10.1007/s10722-003-6017-5

Oh E, Lee MH, Kim JI, Kim S, Pae SB, Ha TJ (2018) Estimation of oil yield of perilla by seed characteristics and crude fat content. Korean J Crop Sci 63(2):158–163. https://doi.org/10.7740/kjcs.2018.63.2.158

Ooijen JW et al. (2006) Joinmap 4: software for the calculation of genetic linkage maps in experimental populations. Kyazma BV, Wageningen 33(10.1371)

Park G, Jang HA, Jo SH, Park Y, Oh SK, Nam M (2018) Development of SNP marker set for marker-assisted backcrossing (MABC) in cultivating tomato varieties. Korean J Agri Sci 45(3):385–400. https://doi.org/10.7774/kjosa.20180061

Poland JA, Brown PJ, Sorrells ME, Jannink JL (2012) Development of high-density genetic maps for barley and wheat using a novel two-enzyme genotyping-by-sequencing approach. PLoS One 7(2):e32253. https://doi.org/10.1371/journal.pone.0032253
Rasheed A, Hao Y, Xia X, Khan A, Xu Y, Varshney RK, He Z (2017) Crop breeding chips and genotyping platforms: progress, challenges, and perspectives. Mol Plant 10(8):1047–1064. https://doi.org/10.1016/j.molp.2017.06.008

Rossetto M, Henry RJ (2014) Escape from the laboratory: new horizons for plant genetics. Trends Plant Sci 19(9):554–555. https://doi.org/10.1016/j.tplants.2014.06.011

Sa KJ, Choi IY, Park KC, Lee JK (2018) Genetic diversity and population structure among accessions of *Perilla frutescens* (L.) britton in east asia using new developed microsatellite markers. Genes Genom 40(12):1319–1329. https://doi.org/10.1007/s13258-018-0727-8

Tamura K, Sakamoto M, Tanizawa Y, Mochizuki T, Matsushita S, Kato Y, Ishikawa T, Okuhara K, Nakamura Y, Bono H (2023) A highly contiguous genome assembly of red perilla (*Perilla frutescens*) domesticated in Japan. DNA Res 30(1):1–8. https://doi.org/10.1093/dnares/dsac044

Verma S, Gupta S, Bandhiwal N, Kumar T, Bharadwaj C, Bhatia S (2015) High-density linkage map construction and mapping of seed trait QTLs in chickpea (*Cicer arietinum* L.) using genotyping-by-sequencing (GBS). Sci Rep 5:17512. https://doi.org/10.1038/srep17512

Voorrips R (2002) Mapchart: software for the graphical presentation of linkage maps and qtls. J Hered 93(1):77–78. https://doi.org/10.1093/jhered/93.1.77

Ye G, Smith KF (2008) Marker-assisted gene pyramiding for inbred line development: basic principles and practical guidelines. Int J Plant Breed 2(1):1–10

**Publisher’s Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.