Identification of Accession-Specific Variants and Development of KASP Markers for Assessing the Genetic Makeup of Crop Seeds

Seongmin Hong
Korea Research Institute of Bioscience and Biotechnology (KRIBB)
Su Ryun Choi
Chungnam National University
Jihyeong Kim
Foundation of Agri. Tech. Commercialization & Transfer (FACT)
Young-Min Jeong
Foundation of Agri. Tech. Commercialization & Transfer (FACT)
Suk-Yoon Kwon
Korea Research Institute of Bioscience and Biotechnology
Yong Pyo Lim
Chungnam National University
Ah-Young Shin
Korea Research Institute of Bioscience and Biotechnology
Yong-Min Kim (ymkim@kribb.re.kr)
Korea Research Institute of Bioscience and Biotechnology (KRIBB)

Research Article

Keywords: Seed Purity Assessment, KASP Marker, Accession-Specific Marker, Crop Breeding, Brassica rapa

Posted Date: December 29th, 2021

DOI: https://doi.org/10.21203/rs.3.rs-1119082/v1

Abstract

Background

Most crop seeds are F1 hybrids. Seed providers and plant breeders must be confident that the seed supplied to growers is of known, and uniform, genetic makeup. This requires maintenance of pure genotypes of the parental lines and testing to ensure the genetic purity of the F1 seed. Traditionally, seed testing for purity was done with a grow-out test (GOT) in the field, but these tests are time consuming and costly. Seed testing with molecular markers was introduced as a replacement for GOT early in the last decade. Recently, Kompetitive allele specific PCR (KASP) markers are promising tools for genetic testing of seeds. However, the markers available at that time could be inaccurate and could be used with only a small number of accessions or varieties due to the limited genetic information and reference genomes available.

Results

Here, we identified 4,925,742 SNPs in 50 accessions of the *Brasscia rapa* core collection. Furthermore, the total 2,925 SNPs were selected as accession-specific SNPs, considering properties of flanking region harboring accession-specific SNPs and genic region conservation among accessions by NGS analysis. In total, 100 accession-specific markers were developed as accession-specific KASP markers. Based on the results of our validation experiments, the accession-specific markers successfully distinguish individuals from the mixed population including 50 target accessions from *B. rapa* core collection and outgroup.

Conclusions

This study provides efficient methods for developing KASP markers to distinguish individuals from the mixture comprised of breeding lines and germplasms from the resequencing data of Chinese cabbage (*Brassica rapa* spp. *pekinensis*).

Background

Most growers of vegetable crops rely on F1 hybrid seeds, and suppliers of these seeds must maintain genetically pure stocks. Not only do the suppliers need to keep seeds of known genetic makeup for sales but also for their ongoing breeding programs. Until the late 1990s, seed providers relied on what is known as the grow-out test (GOT), in which the seeds were planted in the field and the traits of the test plants were assessed by investigation [1]. However, this method is time consuming, requires a large amount of land, and is partly subjective as plant phenotype can be affected by the environment [2]. Thus, precise and efficient tools to assess the genetic makeup and purity of hybrid seeds are sought by seed providers.

In response to these limitations of the GOT, various types of molecular markers have been developed and used to characterize the genotypes of crop plants. This endeavor began in the early 1990s and has resulted in the identification of numerous types of markers. These include restriction fragments length polymorphism, amplified fragments length polymorphism, simple sequence length polymorphism, simple sequence repeat (SSR), and sequence tagged site (STS) markers. The PCR-based SSR or STS markers can be rapidly acquired and are easy to assay, and they have been used for crop breeding or assessment of hybrid seeds in rice, maize, pigeon pea, and pepper [1–4]. However, these markers were developed for specific breeding lines or varieties and are not sufficient to assess the purity of hybrid seeds.

Application of molecular markers to a wide range of situations that require accurate assessment of the genetic makeup of a plant must entail investigating genetic variants in the core collections and commercial lines. Previously, investigation of genetic variants of core collections and commercial lines of crops was limited because of the expense of sequencing and the absence of reference genomes. With the advent of next-generation sequencing technology, reference genomes have been constructed for a number of crops, including tomato [5], pepper [6], cucumber [7], melon [8, 9], wheat [10], and Chinese cabbage [11]. Whole-genome resequencing of various crops has also been undertaken. This has allowed the development of widely applicable molecular markers, accomplished by resequencing analyses of core collections. Also, the development of the Kompetitive Allele Specific PCR genotyping (KASP) assay has permitted the development of accession-specific markers for large-scale seed purity assessments [12–14].

Here, we present pipelines for the detection of accession-specific genetic variants and accession-specific markers from 50 Chinese cabbage accessions. The pipelines were constructed with a combination of genetic variants calling, detection of accession-specific variants, and determination KASP marker candidate sequences. Accession-specific single nucleotide polymorphisms (SNPs) were identified from 50 Chinese cabbage core collections, and 100 accession-specific KASP markers from 50 accessions were developed from a pool of these SNPs. Then, evaluation of KASP markers was carried out using the core collection and 35 non-core collections. We have identified 100 KASP markers that we believe will be useful in assessing hybrid seed purity.

Results

Detection and evaluation of accession-specific variants

We performed genome resequencing analysis of 50 accessions from the *Brassica rapa* core collection, with the goal of developing markers specific to each accession. This core collection is composed of four different groups: non-pekinensis, Chinese, Japanese, and Korean breeding lines (**Fig. 1 and**

Supplementary Table 1). The reads from the analysis of these accessions were mapped to the *Brassica rapa* reference genome (ver 3.0) [11] with the BWA-MEM (ver 0.1.17) using the default parameters. We detected a total of 4,925,742 SNPs from the 50 accessions (**Table 1 and Supplementary Data 1**). Our goal was to identify genetic variants from the *B. rapa* core collection. To this end, we constructed a variant-identification pipeline by combining the calling and filtering variants (**Supplementary Figure 1**). First, SNPs of individual accessions were detected and merged in the joint variant calling step. Then, homozygous alternative alleles for single accessions were identified as accession-specific SNPs by comparing the genotype of each individual accession in the core collection. To develop KASP markers, each accession-specific marker was evaluated by considering the non-redundant flanking sequences, overlapping of repeat sequences, and annotation of the SNPs. Finally, SNPs with unique flanking sequences without overlapping repeat sequences were identified as candidates for development of KASP markers. We identified 2,925 accession-specific SNPs as such candidates (**Table 1**). Almost all of these SNPs were in flanking sequences of genes and 2,806 of them, or approximately 95.9%, were in genic regions (**Table 2**). Of the 2,925 SNPs, approximately 456, or 15.6%, resulted in non-synonymous mutations, and 19 variants led to abnormal termination of translation. These genetic variants may be important in future investigation of trait-associated genes or markers. Our next step in the development of accession-specific markers was to validate the SNPs with genome resequencing analysis, which we did with Sanger sequencing (**Fig. 2**).

Eight flanking sequences of the accession-specific SNP candidates were selected from the four groups of the core collection. Primers for Sanger sequencing were designed (**Supplementary Table 2**). From the results of the Sanger sequencing, we concluded that 7 of the SNP candidates were specific to a single accession (**Fig. 2 and Supplementary Figure 2-7**). Amplification by PCR for Sanger sequencing failed in one flanking sequences (**Supplementary Figure 8**), leading us to conclude that SNPs with conserved flanking sequences were the best candidates for developing accession-specific markers with PCR. Also, candidate SNPs with highly conserved flanking sequences that are suitable for primers may be necessary for developing wide-ranging KASP markers that will apply to crops not in the core collection or to commercial cultivars. Clearly, determination of primer sites for KASP markers is important for the development of accession-specific KASP markers.

Development and evaluation of KASP markers

Our next venture was to develop accession-specific KASP markers for assessment of hybrid seed purity. Five of the accession-specific SNP candidates identified as described above were selected from individual accessions for further analysis. Primer sites played an important role in successful marker development, and conserved flanking sequences of SNPs in our core collections were surveyed (**Fig. 3a**). Flanking regions containing non-sequence sites, shown as N in the reference genome, were removed from the primer candidate sequences (**Fig. 3b**). Then, five flanking sequences in each accession-specific SNP were selected for further evaluation of KASP markers. It was necessary to consider the genomic position of the SNP in the development of a wide range of markers, as overlapping genomic positions among markers may lead to inefficiency or false positive results when seed purity is assessed. To avoid this redundancy, the genomic positions of five candidate SNPs from individual accessions were investigated and the positions unique to the accessions were selected (**Fig. 4**). In total, two SNPs in each accession were selected for validation of KASP markers (**Supplementary Table 3)**. Many of the KASP markers that were in genic regions caused non-synonymous variation, although almost all accession-specific SNPs were detected in the flanking regions of genes (**Table 2**).

Validation of KASP markers was carried out using 50 accessions from core collection and 35 from non-core collections or commercial cultivars to determine their applicability to a wide range of seed purity assessments (**Fig. 5, Table 3, and SupplementaryData2**). Based on the results, we conclude that accession-specific markers were successfully distinguished in individual accessions in both the core collection and the outgroup (**Fig. 5**). We suggest that accession-specific markers developed using a large amount of individual resequencing data can be used to assess seed purity of seed from of non-sequenced accessions or cultivars. The accession-specific markers developed here should be useful in a wide range of seed purity assessments in crop breeding and commercial seed production.

Discussion

Crop breeders need to maintain accessions and varieties of crop plants of known genetic makeup for the successful development of new varieties. Seed purity is traditionally estimated with the GOT in the field [1], but this test is both time consuming and expensive. The use of molecular markers to identify the genotype of seeds and plant material holds promise for replacing the GOT. Although the markers are faster and less expensive that the GOT, at present, they can be inaccurate and most of them were developed early in the 2010s for only a small number of accessions or varieties. With the advent of next-generation sequencing technology, construction of high-quality reference genomes and genetic information for many different cultivars and species has been generated. This information should provide the background necessary for the development of molecular markers that will provide accurate information and will be useful in a wide range of applications. These will include studying genetic variants in individual accessions, varieties, and large populations [25]. Reference genomes also provide useful detailed information on genetic variants such as gene structures, repetitive sequences, and accurate positions of various genetic features. This technology can also be applied to correlation analyses of phenotypes and may prove useful in analyses such as quantitative trait locus mapping and genome-wide association studies (GWAS) [26–28].

In the current study, we identified SNPs in the *B. rapa* core collection with genome resequencing (Fig. 1). Examining accession-specific genetic variants, we identified 4,925,742 SNPs in 50 accessions and, among these, we identified 2,925 SNPs that were specific to a single accession (Table 1). Almost all of the genetic variants we detected were in flanking regions of genes, but KASP markers were developed from SNPs that caused non-synonymous variations and were in genic regions. Conservation of the genic regions could maintain the function of the genes, accounting for our observation that the

ratio of conserved sequences was relatively greater than for the other regions. The non-synonymous mutations might be involved in phenotypic or morphological differences among accessions and should be useful in investigation of trait-associated genes or markers associated with traits.

Until quite recently, molecular markers have not been developed for crops or cultivars, and their application has been limited. Our development of molecular markers using the core collection of *B. rapa* was performed, in part, to address this problem: we sought to develop markers, considering conserved sequence for primer sites, for a wide range of applications. (Fig. 3). Furthermore, genomic positions of accession-specific markers were investigated to avoid overlapping of the genomic positions of KASP markers (Fig. 4). In total, 100 accession-specific markers were developed as accession-specific KASP markers. Based on the results of our validation experiments, we conclude that the accession-specific markers were successfully distinguished in individual accessions in test populations from non-core or commercial cultivars (Fig. 5).

Conclusions

This study shows efficient methods for developing KASP markers to distinguish individuals from the mixture comprised of breeding lines and germplasms from the resequencing data of Chinese cabbage (*Brassica rapa* spp. *pekinensis*). We show that the accession-specific SNPs identified by NGS data pipelines are feasible targets to develop KASP markers. We believe the KASP markers developed here will be applicable to assessment of seed purity in a wide variety of situations, including core collections, other non-sequenced accessions, or commercial cultivars. These markers should also prove useful to breeding programs of *B. rapa*, facilitating the essential maintenance of pure parental lines. Furthermore, the non-synonymous mutations detected here should aid investigations of genes or markers associated with traits and in functional studies of genes. This study will help marker development to check the seed purity of commercial F1 seed samples whether they are produced by unintended crossing or not.

Methods

Plant materials

To develop accession-specific KASP markers, 50 accessions of *Brassica rapa* core collections [15] were used for whole-genome resequencing analysis. These accessions were characterized as inbred lines or doubled haploid lines. For assessments of KASP markers, 35 accessions (F1 hybrids and germplasm) donated by Chungnam National University (CNU) and Rural Development Administration were used as the control panel for validation of the KASP markers.

Genome resequencing of core collection

Truseq Nano DNA libraries were constructed according to the manufacturer's instructions. In total, 100 ng or 200 ng of high molecular weight genomic DNA to generate a large (550 bp) insert size were sheared to yield DNA fragments using Covaris S2 system. Blunt-ended DNA fragments were generated with a combination of fill-in reactions and exonuclease activity. A single A-base was then added to the blunt ends of each strand in preparation for ligation to the indexed adapters. Each adapter contained a single T-base overhang for ligating the adapter to the A-tailed fragmented DNA. Ligated products were amplified with reduced-bias PCR. The quality of the amplified libraries was verified with capillary electrophoresis (Bioanalyzer, Agilent). After QPCR using SYBR Green PCR Master Mix (Applied Biosystems), we combined index-tagged libraries in equimolar amounts in the pool. Whole-genome resequencing was performed with an Illumina NovaSeq 6000 system, following the protocols provided for 2×100 sequencing.

Identification of genetic variants

The FastQC (v.0.11.3) program was used to assess quality and to detect adaptor sequences of reads (https://www.bioinformatics.babraham.ac.uk/projects/fastqc/). Adaptor sequences and low-quality reads were filtered using Trimmomatic (ver 0.36) with the parameter ILLUMINACLIP:TruSeq3-PE-2.fa:2:30:10 SLIDINGWINDOW:4:20 TRAILING:20 MINLEN:75 [16]. Then, the filtered reads were aligned to the *Brassica rapa* reference genome (ver 3.0) [11] with Burrows Wheel Aligner (BWA) (ver 0.1.17), using the default parameter [17]. These results (*.sam) were converted to bam files using SAMtools (ver 1.9) [18] and low-quality reads (mapping quality < 30) were removed. Reads duplicated by PCR were also removed, with MarkDuplicate in Picard tools (ver 2.21.1) (http://broadinstitute.github.io/picard/). To detect InDels, InDels of reference genomes were detected using RealignerTargetCreator in GATK (ver 3.7) [19] and reads mapped InDels were re-aligned using IndelRealigner. SNPs were detected and filtered (read depth > 3, genotype quality > 30, homozygous allele only) by BCFtools (ver 1.9) [20]. To determine possible SNP positions in the core collection, joint variant calling was conducted for all possible SNP positions in each accession. Multiple allelic positions and low-depth genotypes (read depth < 3) were filtered by VCFtools (ver 0.1.13) [21].

Construction of a pipeline for accession-specific variants calling

Positions of SNPs that had homozygous alternative alleles for one accession in the population variant call format (vcf) file were selected as accession-specific variants by in-house perl script. To reduce the possibility of primer amplification for multiple loci, target sequence redundancy in the *B. rapa* genome was estimated with megablast task of BlastN [22] with 501 bp sequences harboring accession-specific SNPs. Accession-specific SNPs without flanking sequence redundancy were selected for KASP primer design. Also, accession-specific SNPs with flanking sequence overlapping predicted repeat

sequences were filtered out by utilizing gff file provided by *B. rapa* reference genome ver 3.0. Accession-specific variants on the exon region were given priority for KASP primer design after SNP annotation by snpEFF [23]. A total of 20 accession SNP positions from 10 accessions were confirmed by Sanger sequencing.

Construction of pipeline for KASP marker development

We sought to minimize the failure of primer amplification due to insertion or deletion on the marker target sites **(Fig. 3a)**. This led us to develop a pipeline for producing KASP candidate sequences for accession-specific variants. The pipeline we developed generates flanking region sequences that harbor accession-specific variants from bam files of each accessions and aligns them based on the reference genome sequence with ClustalW (-OUTPUT=CLUSTAL -TYPE=DNA -GAPOPEN=10 -ENDGAPS -GAPDIST=0.05) [24]. The pipeline evaluates the proportion of missing or alternative alleles from all of the aligned positions and produces consensus sequences masking variable positions (non-reference allele for positions > 10%) with N (**Fig. 3b**). Accession-specific variants located at bp 251 on the consensus sequences were directly used for the KASP primer designed by the LGC Genomics.

Evaluation and application of KASP markers

Validation of the KASP markers was performed with the Nexar system (LGC Douglas Scientific, Alexandria, USA) at the Seed Industry Promotion Center of the Foundation of Agricultural Technology Commercialization and Transfer (Gimje, Korea). An aliquot (0.8 L) of 2x Master mix, 0.02 L of 72x KASP assay mix (both from LGC Genomics), and 5 ng genomic DNA template from the 85 *B. rapa* accessions were mixed into 1.6 L of KASP reaction mixture in a 384-well Array Tape. Duplicate reactions were run, and non-template controls were included in each run. KASP amplification was performed with the following thermal cycling profile: 15 min at 94℃, a touchdown phase of 10 cycles at 94℃ for 20 s and at 61℃-55℃, in which the temperature decreased by 0.6℃ per cycle, for 60 s, and 26 cycles at 94℃ for 20 s and 55℃ for 60 s (first PCR stage). Next, recycling was performed with three cycles of 94℃ for 20 s and 57℃ for 60 s (second PCR stage). The recycling was performed twice, and the fluorescence read was taken for KASP genotyping after PCR amplification.

Declarations

Ethics approval and consent to participate

Not applicable

Consent for publication

Not applicable

Availability of data and materials

The datasets are deposited at NCBI under BioProject number PRJNA787013. Genetic variants information (BAM files) are available through the NCBI Sequence Read Archive (SRA) with identifier SRPXXXXXX (under processing).

Competing interests

The authors declare that they have no competing interests

Funding

This work was supported by the Korea Institute of Planning and Evaluation for Technology in Food, Agriculture, and Forestry through the Golden Seed Project (213006-05-5-SBG30) and the Technology Commercialization Support Program (821026-03) funded by a grant from the Ministry of Agriculture, Food, and Rural Affairs, the Ministry of Oceans and Fisheries, the Rural Development Administration, and the Korea Forest Service supported to Y.-M.K. The latter is funded by the Korea Forest Service of the Korean government through its Research and Development Program for Forestry Technology (Project No. 2014071H10-2122-AA04) supported to Y.-M.K.

Authors Contributions

Y.-M.K. conceived the project, designed the experiments, and organized the manuscript. S.R.C. and Y.P.L. generated and maintained the core collection. S.H., A.-Y.S., S.Y.K., Y.P.L., and Y.-M.K. performed the bioinformatics analysis and development of the KASP markers. J.K and Y.-M.J. performed validation of KASP markers. A.-Y.S. and Y.-M.K. wrote the manuscript.

Acknowledgement

Not applicable

References

1. Ballester J, de Vicente MC: **Determination of F1 hybrid seed purity in pepper using PCR-based markers**. *Euphytica* 1998, **103**(2):223-226.
2. Saxena RK, Saxena K, Varshney RK: **Application of SSR markers for molecular characterization of hybrid parents and purity assessment of ICPH 2438 hybrid of pigeonpea [Cajanus cajan (L.) Millspaugh]**. *Molecular Breeding* 2010, **26**(2):371-380.
3. Sundaram RM, Naveenkumar B, Biradar SK, Balachandran SM, Mishra B, IlyasAhmed M, Viraktamath BC, Ramesha MS, Sarma NP: **Identification of informative SSR markers capable of distinguishing hybrid rice parental lines and their utilization in seed purity assessment**. *Euphytica* 2008, **163**(2):215-224.
4. Komori T, Nitta N: **A simple method to control the seed purity of japonica hybrid rice varieties using PCR-based markers**. *Plant Breeding* 2004, **123**(6):549-553.
5. Tomato Genome C: **The tomato genome sequence provides insights into fleshy fruit evolution**. *Nature* 2012, **485**(7400):635-641.
6. Kim S, Park M, Yeom SI, Kim YM, Lee JM, Lee HA, Seo E, Choi J, Cheong K, Kim KT *et al*: **Genome sequence of the hot pepper provides insights into the evolution of pungency in Capsicum species**. *Nat Genet* 2014, **46**(3):270-278.
7. 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**(6).
8. Garcia-Mas J, Benjak A, Sanseverino W, Bourgeois M, Mir G, Gonzalez VM, Henaff E, Camara F, Cozzuto L, Lowy E *et al*: **The genome of melon (Cucumis melo L.)**. *Proc Natl Acad Sci U S A* 2012, **109**(29):11872-11877.
9. Shin AY, Koo N, Kim S, Sim YM, Choi D, Kim YM, Kwon SY: **Draft genome sequences of two oriental melons, Cucumis melo L. var. makuwa**. *Sci Data* 2019, **6**(1):220.
10. International Wheat Genome Sequencing C, investigators IRp, Appels R, Eversole K, Feuillet C, Keller B, Rogers J, Stein N, investigators Iw-gap, Pozniak CJ *et al*: **Shifting the limits in wheat research and breeding using a fully annotated reference genome**. *Science* 2018, **361**(6403).
11. Zhang L, Cai X, Wu J, Liu M, Grob S, Cheng F, Liang J, Cai C, Liu Z, Liu B *et al*: **Improved Brassica rapa reference genome by single-molecule sequencing and chromosome conformation capture technologies**. *Hortic Res* 2018, **5**:50.
12. Semagn K, Babu R, Hearne S, Olsen M: **Single nucleotide polymorphism genotyping using Kompetitive Allele Specific PCR (KASP): overview of the technology and its application in crop improvement**. *Molecular Breeding* 2014, **33**(1):1-14.
13. Rasheed A, Wen W, Gao F, Zhai S, Jin H, Liu J, Guo Q, Zhang Y, Dreisigacker S, Xia X *et al*: **Development and validation of KASP assays for genes underpinning key economic traits in bread wheat**. *Theor Appl Genet* 2016, **129**(10):1843-1860.
14. Jagtap AB, Vikal Y, Johal GS: **Genome-Wide Development and Validation of Cost-Effective KASP Marker Assays for Genetic Dissection of Heat Stress Tolerance in Maize**. *International journal of molecular sciences* 2020, **21**(19).
15. Pang W, Li X, Choi SR, Dhandapani V, Im S, Park MY, Jang CS, Yang M-S, Ham IK, Lee EM *et al*: **Development of a leafy Brassica rapa fixed line collection for genetic diversity and population structure analysis**. *Molecular Breeding* 2015, **35**(1):54.
16. Bolger AM, Lohse M, Usadel B: **Trimmomatic: a flexible trimmer for Illumina sequence data**. *Bioinformatics* 2014, **30**(15):2114-2120.
17. Li H, Durbin R: **Fast and accurate short read alignment with Burrows-Wheeler transform**. *Bioinformatics* 2009, **25**(14):1754-1760.
18. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, Genome Project Data Processing S: **The Sequence Alignment/Map format and SAMtools**. *Bioinformatics* 2009, **25**(16):2078-2079.
19. McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky 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 research* 2010, **20**(9):1297-1303.
20. Narasimhan V, Danecek P, Scally A, Xue Y, Tyler-Smith C, Durbin R: **BCFtools/RoH: a hidden Markov model approach for detecting autozygosity from next-generation sequencing data**. *Bioinformatics* 2016, **32**(11):1749-1751.
21. Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, Handsaker RE, Lunter G, Marth GT, Sherry ST *et al*: **The variant call format and VCFtools**. *Bioinformatics* 2011, **27**(15):2156-2158.
22. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ: **Basic local alignment search tool**. *J Mol Biol* 1990, **215**(3):403-410.
23. Cingolani P, Platts A, Wang le L, Coon M, Nguyen T, Wang L, Land SJ, Lu X, Ruden DM: **A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of Drosophila melanogaster strain w1118; iso-2; iso-3**. *Fly (Austin)* 2012, **6**(2):80-92.
24. Thompson JD, Higgins DG, Gibson TJ: **CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice**. *Nucleic Acids Res* 1994, **22**(22):4673-4680.
25. Chen Q, Song J, Du WP, Xu LY, Jiang Y, Zhang J, Xiang XL, Yu GR: **Identification, Mapping, and Molecular Marker Development for Rgsr8.1: A New Quantitative Trait Locus Conferring Resistance to Gibberella Stalk Rot in Maize (Zea mays L.)**. *Frontiers in plant science* 2017, **8**:1355.
26. Du X, Huang G, He S, Yang Z, Sun G, Ma X, Li N, Zhang X, Sun J, Liu M *et al*: **Resequencing of 243 diploid cotton accessions based on an updated A genome identifies the genetic basis of key agronomic traits**. *Nat Genet* 2018, **50**(6):796-802.
27. Lu K, Wei L, Li X, Wang Y, Wu J, Liu M, Zhang C, Chen Z, Xiao Z, Jian H *et al*: **Whole-genome resequencing reveals Brassica napus origin and genetic loci involved in its improvement**. *Nat Commun* 2019, **10**(1):1154.
28. Yuan J, Wang X, Zhao Y, Khan NU, Zhao Z, Zhang Y, Wen X, Tang F, Wang F, Li Z: **Genetic basis and identification of candidate genes for salt tolerance in rice by GWAS**. *Scientific reports* 2020, **10**(1):9958.

Tables

Table 1. The single nucleotide polymorphisms (SNPs) that were identified from 50 *B. rapa* accessions.

Group	Numbers of accessions	Total number of identified SNPs	Total number of accession-specific SNPs
Korean	34	4,095,628	1,314
Chinese	5	3,289,533	357
Japanese	3	2,191,837	325
Non-pekinensis	2	2,016,767	429
Others	6	3,044,951	500
Total	50	4,925,742	2,925

Table 2. Annotation of the accession-specific single nucleotide polymorphisms (SNPs) that were identified from the *B. rapa* core collection.

Annotation of variants	Type	Accession-specific SNPs (No.)	KASP markers (No.)
Variant causes a codon that produces a different amino acid	Exon	456	72
Variant causes a codon that produces the same amino acid	Exon	189	18
Variant causes a STOP codon	Exon	17	2
Variant causes start codon to be mutated into a non-start codon	Exon	1	0
Variant causes stop codon to be mutated into a non-stop codon	Exon	1	0
Variant causes stop codon to be mutated into another stop codon	Exon	3	1
The variant hits a splice acceptor site	Intron	5	0
The variant hits a Splice donor site	Intron	4	0
Variant hits intron	Intron	122	1
Downstream of a gene (default length: 5K bases)	Non-coding	459	0
The variant is in an intergenic region	Non-coding	119	0
Upstream of a gene (default length: 5K bases)	Non-coding	1549	6
Total	-	2,925	100

Table 3. Accession-specific single nucleotide polymorphisms (SNPs) that were identified in eight representative *B. rapa* accessions. (Reference genome, REF; allele of reference genome and genetic variants, ALT; Chungnam National University, CNU, a provider of 35 accessions used in this study)

Category	Accession ID	Chromosome	Position	REF	ALT	CNU_11479	CNU_11480	26021	26022	28059	28061	CNU_11411	CNU_11412
Non-pekinensis	CNU_11479	A03	21,060,850	A	C	CC	AA	AA	AA	AA	AA	AA	AA
		A06	3,079,806	A	G	GG	AA	AA	AA	AA	AA	AA	AA
	CNU_11480	A02	15,635,917	C	T	CC	TT	CC	CC	CC	CC	CC	CC
		A03	10,121,108	G	A	GG	AA	GG	GG	GG	GG	GG	GG
Chinese germplasm	26021	A02	18,369,509	C	T	CC	CC	TT	CC	CC	CC	CC	CC
		A06	6,118,599	G	A	GG	GG	AA	GG	GG	GG	GG	GG
	26022	A01	6,548,614	A	T	AA	AA	AA	TT	AA	AA	AA	AA
		A03	3,737,651	A	C	AA	AA	AA	CC	AA	AA	AA	AA
Japanese breeding	28059	A07	21,078,330	G	T	GG	GG	GG	GG	TT	GG	GG	GG
		A07	23,181,319	G	T	GG	GG	GG	GG	TT	GG	GG	GG
	28061	A03	22,285,257	G	A	GG	GG	GG	GG	GG	AA	GG	GG
		A10	450,622	A	C	AA	AA	AA	AA	AA	CC	AA	AA
Korean breeding	CNU_11411	A07	20,012,970	G	A	GG	GG	GG	GG	GG	GG	AA	GG
		A09	37,233,481	C	T	CC	CC	CC	CC	CC	CC	TT	CC
	CNU_11412	A07	21,781,162	C	G	CC	CC	CC	CC	CC	CC	CC	GG
		A09	42,427,036	C	G	CC	CC	CC	CC	CC	CC	CC	GG

Figures

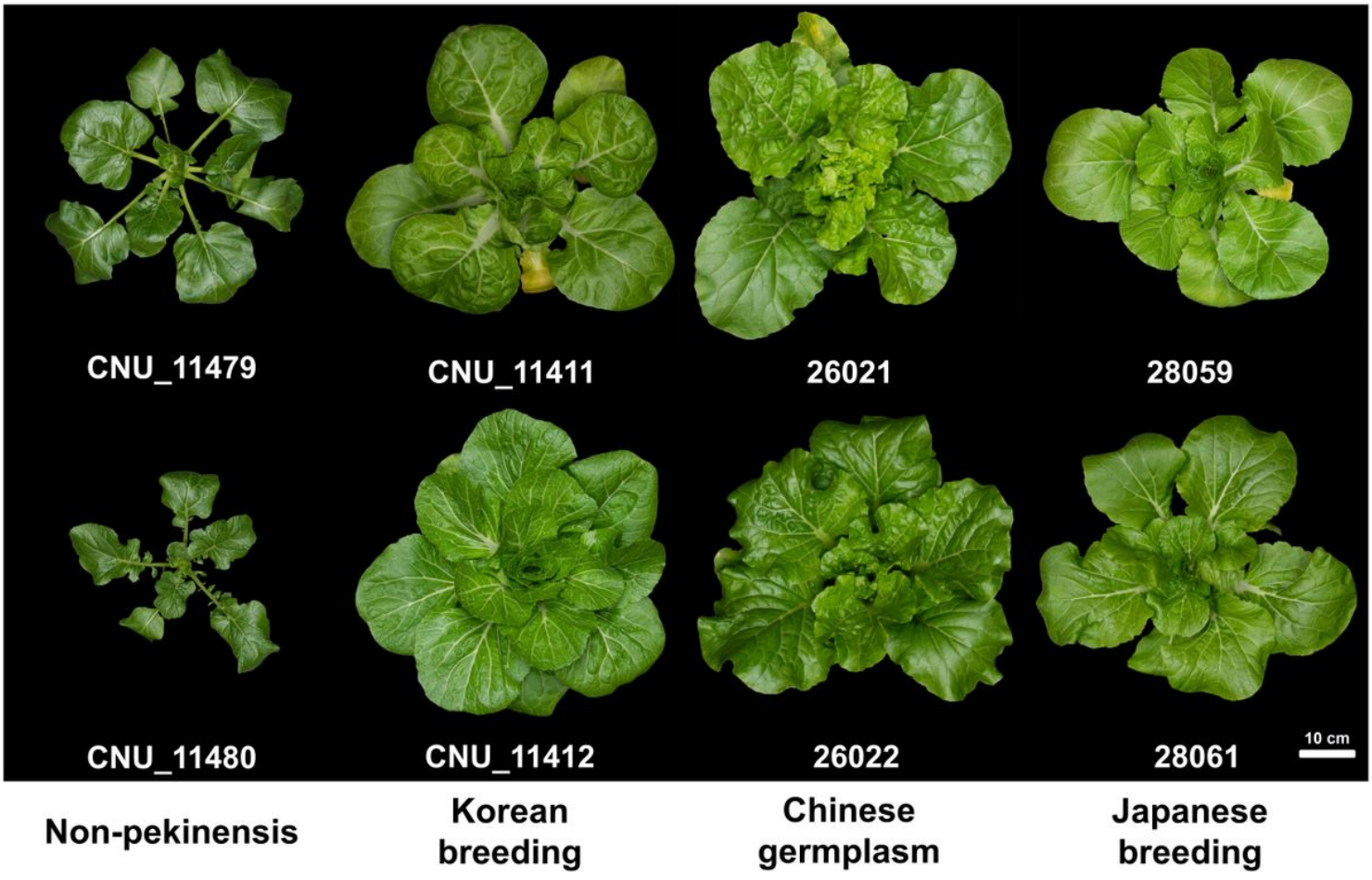

Figure 1

Morphological features of eight representative accessions from four groups of the *Brassica rapa* core collection.

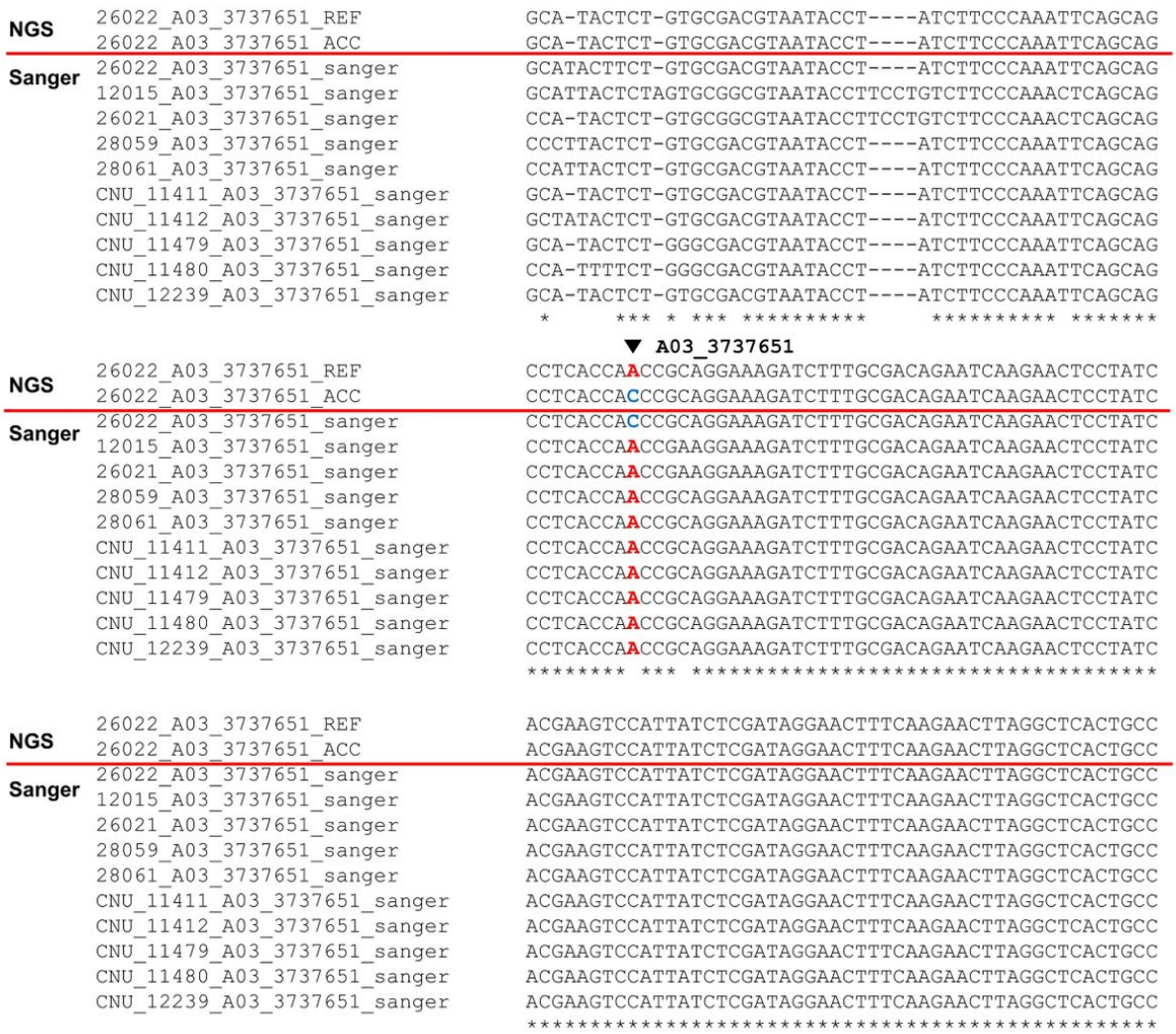

Figure 2

Validation of accession-specific single nucleotide polymorphisms (SNPs) (3,737,651 in chromosome 3) from accession 26022 (from Chungnam National University) using *the Brassica rapa* reference genome (ver 3.0). (REF, reference genome; ACC, resequencing result of individual accession; Sanger, Sanger sequencing result)

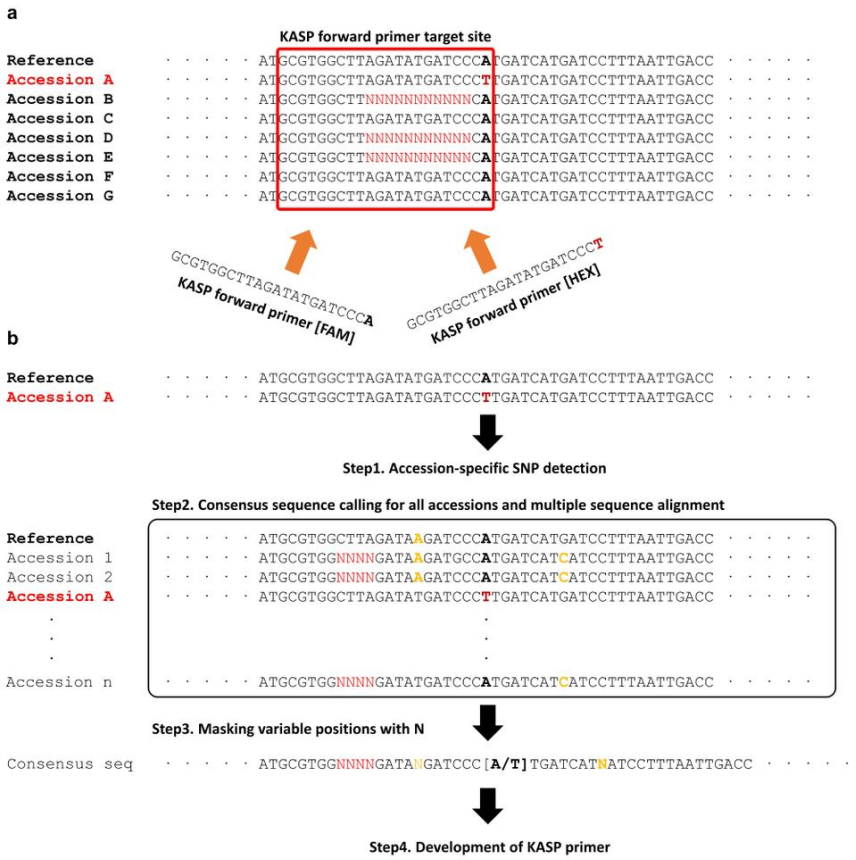

Figure 3

Development of KASP markers. a) Potential problem of primer alignments by possible sequence variation from core collection during KASP marker development, b) Process for development of KASP markers.

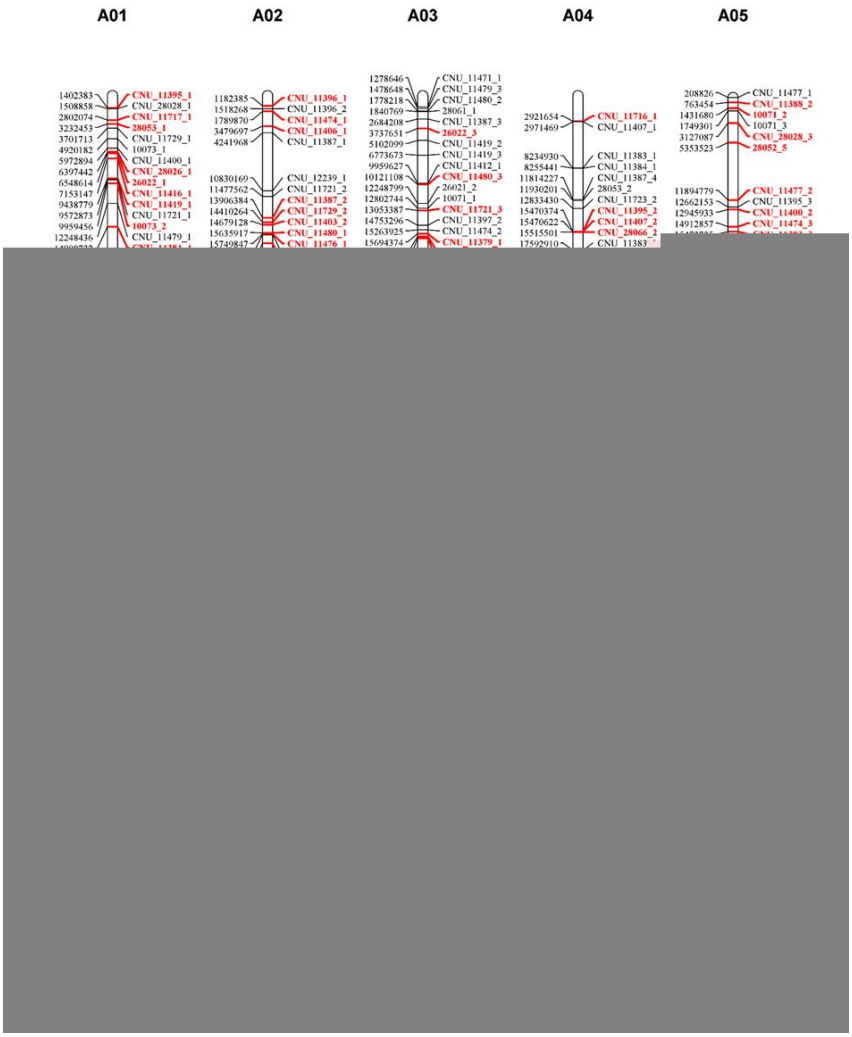

Figure 4

Genome distribution of accession-specific SNPs from the *Brassica rapa* core collection. The genomic positions of five accession-specific SNPs in each accession were investigated to develop KASP markers. (The marker positions with red color stand for SNPs used for KASP makrer development)

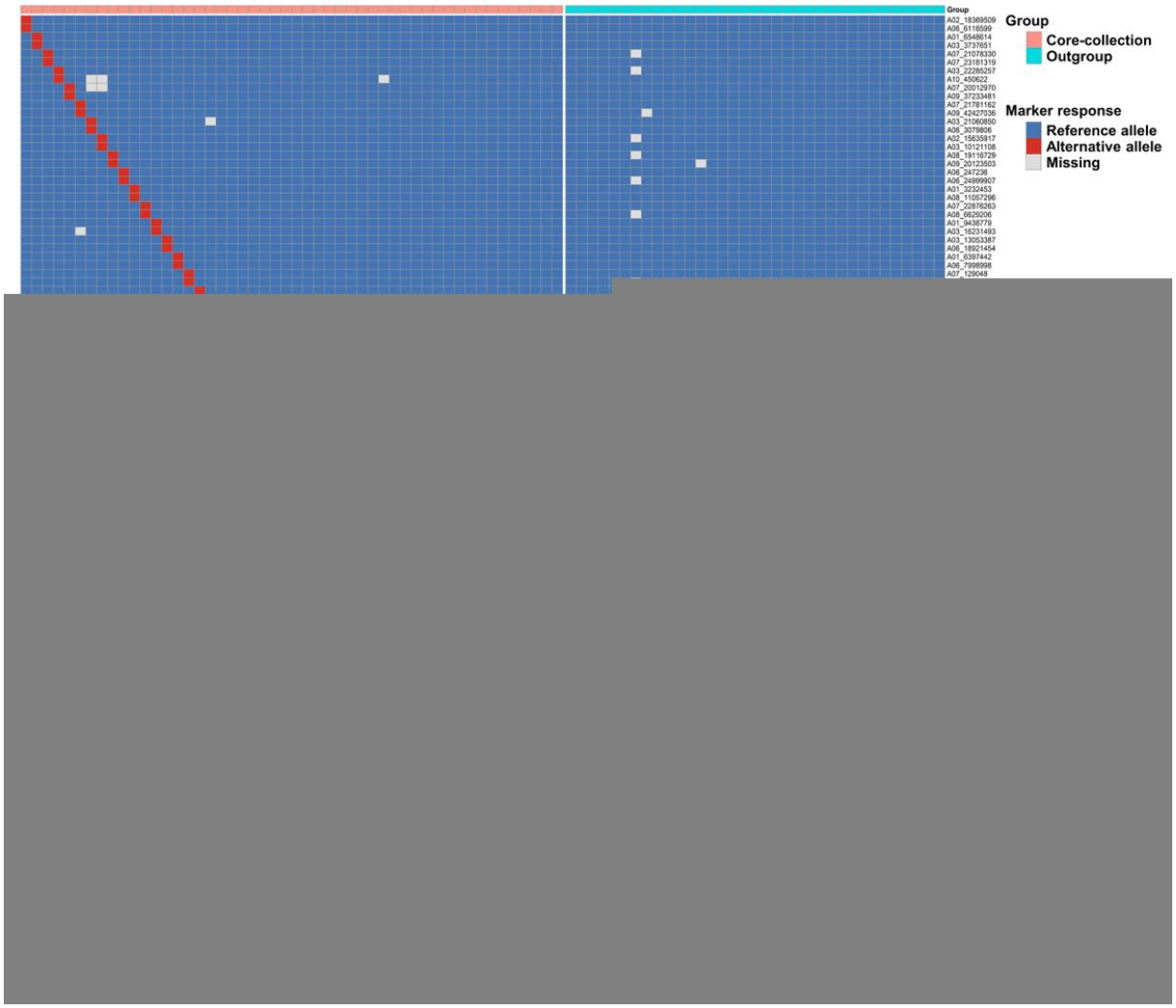

Figure 5

Validation of KASP markers using the *Brassica rapa* core collection, non-core collection, and commercial varieties. (Red bar on the top of heatmap stands for core collection, and blue bar stands for outgroup.)

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- Supplementarydata.xlsx
- Supplementaryfigure.docx
- Supplementarytable.xlsx