Preprints.org

Article

Construction and Characterization of an F1DH Population and High-Density Genetic Map for Ornamental Kale Breeding

Ning Guo [*] , Fan Liu [*] , Shuo Han , Mei Zong , Guixiang Wang , Mengmeng Duan

Article

Construction and Characterization of an F1DH Population and High-Density Genetic Map for Ornamental Kale Breeding

Ning Guo *, Shuo Han, Mei Zong, Guixiang Wang, Mengmeng Duan and Fan Liu *

State Key Laboratory of Vegetable Biobreeding, National Engineering Research Center for Vegetables, Beijing Key Laboratory of Vegetable Germplasms Improvement, Key Laboratory of Biology and Genetics Improvement of Horticultural Crops (North China), Beijing Vegetable Research Center, Beijing Academy of Agriculture and Forestry Science, Beijing 100097, China

* Correspondence: guoning@nercv.org (N.G.); liufan@nercv.org (F.L.)

Abstract: Ornamental kale (*Brassica oleracea* var. *acephala*) is an attractive ornamental plant with a range of leaf colors and shapes. Breeding new varieties of ornamental kale has proven challenging due to its lengthy breeding cycle and the limited availability of genetic markers. Microspore culture is an effective approach to generate new materials and genetic maps are a prerequisite for quantitative trait loci analysis, MAS, fine gene mapping, and genome sequence assembly. In this study, a F_1DH ornamental kale population comprising 300 DH lines was constructed using microspore culture. A high-density genetic map was developed by conducting whole-genome sequencing on 150 individuals from the F_1DH population. The genetic map contained 1,696 bin-markers with 982,642 single-nucleotide polymorphisms (SNPs) spanning a total distance of 775.81 cM on all nine chromosomes with an average distance between markers of 0.46 cM. The ornamental kale genetic map contained substantially more SNP markers compared with published genetic maps for other *B. oleracea* crops. The F_1DH progenies provide an excellent resource for germplasm innovation and breeding new varieties of ornamental kale. The high-density genetic map provides crucial insights for gene mapping and unraveling the molecular mechanisms behind important agronomic traits in ornamental kale.

Keywords: ornamental kale; F_1DH population; whole-genome resequencing; high-density genetic linkage map

1. Introduction

Ornamental kale (*Brassica oleracea* var. *acephala* DC.), as a cultivated variety of *Brassica oleracea* (CC, 2n = 18), is an excellent ornamental foliage plant with a range of leaf colors and shapes. It is widely cultivated as a landscaping plant, potted plant, and for cut foliage. The plants are sufficiently tolerant of frost and chilling that they can grow vigorously in regions that experience low temperatures [1]. The ornamental value and cold tolerance make it a desirable bedding plant in cold seasons and areas. Compared with other *Brassica* crops, fewer varieties of ornamental kale are available commercially and the availability of breeding resources is limited, which has greatly hindered germplasm innovation and the breeding of new varieties.

Brassica oleracea shows obvious heterosis, and cross-breeding is used to produce F_1 hybrids, which ensures strong uniformity and protection of varietal rights, and is the main method used to breed new ornamental kale varieties. Inbred (true-breeding) lines that are homozygous at virtually all loci enable both consistent production of superior hybrid plants as well as genetic analysis [2–4]. Traditional inbreeding, however, requires 6–8 generations of selfing or sib-mating to generate inbred lines, which is labor-intensive and time-consuming. However, the recent technique of isolated microspore culture offers a faster way to produce doubled-haploid (DH) plants, which are essential for consistent hybrid production and genetic analysis [5,6]. Consequently, this technique has been

successfully applied to morphotypes of *B. oleracea*, such as cabbage, broccoli and kale, and large-scale DH lines have been developed and used in breeding [6–8].

Populations of DH lines have been used extensively for high-density genetic linkage map construction and quantitative trait locus (QTL) mapping, as they are permanent and completely homozygous. DH lines have clear advantages over F_2 populations, which consist of individual plants that are often heterozygous. However, the rate of recombination between the parental genomes is lower in DH lines compared to recombinant inbred lines. Creating and maintaining recombinant inbred lines is challenging due to the self-incompatibility of most *B. oleracea* accessions [9]. A DH population provides many homozygous progenies with abundant phenotypic diversity suitable as breeding resources.

Agronomic traits, which are vital for crop improvement, are often complex. Genetic maps, built over the past 30 years for *B. oleracea*, serve as a foundation for studies like QTL mapping and gene fine mapping. Considering all these genetic linkage maps, molecular markers ranging in number from 92 to 4,787 were used and the linkage group lengths ranged from 65 to 1,738 cM [10]. The first high-density genetic linkage map of *B. oleracea*, with 1,257 molecular markers including sequence-related amplified polymorphism and simple sequence repeat (SSR) markers spanning 703 cM in nine linkage groups, was constructed based on a F_2 population derived from a broccoli and cauliflower cross [11]. The second high-density genetic linkage map was constructed including 1,227 markers (602 SSR and 625 single-nucleotide polymorphism (SNP) markers) in nine linkage groups spanning a total of 1197.9 cM with an average of 0.98 cM, using a cabbage DH population. This map also allowed the assembled scaffolds to be anchored to pseudochromosomes [12]. Next-generation sequencing (NGS) technologies have enabled the development of genome-wide methodologies for construction of ultra-high-density genetic linkage maps in different crops, thus allowing placement of candidate loci within several kilobases in a genome [13]. With advances in sequencing technologies, several reference genomes have been generated for different morphotypes of *B. oleracea*, including kale [14], cabbage [15–18], cauliflower [18,19], and broccoli [20], which greatly facilitates development of a large number of SNP markers for genetic map construction, leading to the improved efficiency of fine mapping. A high-density genetic linkage map of cabbage was constructed with 4,103 genotyping-by-sequencing SNP markers in $F_{2:3}$ progenies; this map covered a total genetic distance of 879.9 cM [21]. A genetic map of cauliflower was constructed with 1,776 specific locus amplified fragment (SLAF) markers spanning a total genetic length of 890.01 cM with an average marker interval of 0.50 cM [22]. Recently, a genetic map was constructed with 4,787 SLAF markers with a mean marker distance of 0.22 cM in a DH population of broccoli, and loci controlling the hollow stem trait were identified in the genetic map [23]. No high-density linkage genetic maps of ornamental kale have been reported, which limits gene mapping and molecular breeding for agronomic traits to some extent.

To generate new breeding materials and map the genes that control important agronomic traits of ornamental kale, an ornamental kale F_1-derived doubled-haploid (F_1DH) population derived from parents that differed markedly in leaf shape and color was constructed using microspore culture. Based on this permanent genetic population, a high-density genetic linkage map was constructed with 1,696 bin-markers containing 982,642 SNPs by high-throughput whole-genome resequencing. This map covered a total genetic distance of 775.81 cM in nine linkage groups with an average marker interval of 0.46 cM. This research will lay a foundation for proliferation of ornamental kale breeding and the cloning of agronomically important genes.

2. Materials and Methods

2.1. Parental materials and F_1 progeny used to generate the F_1DH population

Two DH lines, 05-DH-65 (red inner leaves, round leaf, cabbage type) and 06-DH-71 (white inner leaves, lobed leaf, feathered type), which were derived from the commercial cultivars "Paradise Red" and "Coral Sea White" by microspore culture respectively, were selected as parents to generate the F_1 hybrid (pink inner leaves, lobed leaf) (Figure 1).

2.2. Development of the F_1DH population using microspore culture

We conducted microspore culture to obtain DH lines based on a published procedure with minor modifications [24–26]. Inflorescences with microspores in the late uninucleate stage were selected for microspore isolation. Buds were excised and subjected to surface sterilization by dipping briefly in 70% ethanol and then soaking in 2% NaOCl for 10 minutes. Subsequently, the buds were rinsed three times with sterile distilled water for durations of 1, 4, and 10 minutes, respectively. The buds were then gently macerated in a tube containing NLN13 liquid medium. The microspores were collected by filtration through two layers of 44 μm nylon filters and subsequently washed three times by centrifugation at 100 g for 3-4 minutes each time. After the final wash, the microspores were resuspended in NLN13 medium at a density of 4×10^4 ml^{-1}, and the suspension was poured into Petri dishes for culturing. The cultures were incubated in the dark at 31°C for 72 hours, after which they were transferred to 25°C, again in the dark. Finally, two to eight embryos from each Petri dish were germinated and rooted on B5-Gamborg medium containing 1% sucrose. Well-rooted plants were then transferred to the glasshouse [26].

2.3. Flow cytometry analysis

All plants regenerated from microspore culture were analyzed by flow cytometry to identify the ploidy level. About 1.5 cm^2 young leaf tissue was chopped in nucleus extraction buffer for suspension of intact nuclei [27], and stained with propidium iodide (50 μg/ml) in the dark for 30 min. The samples were analyzed with a flow cytometer (FACSCalibur, BD Biosciences, San Jose, CA, USA) to determine the ploidy. The relative DNA content of the DH parents was adjusted to "channel 100" as the standard. Each histogram was generated from the analysis of at least 10,000 cells for each of four replicates.

2.4. Plant materials for genetic map construction and DNA extraction

The 150 F_1DH lines were grown together with the parental lines in the greenhouse of the Beijing Vegetable Research Center (Beijing, China). Young fresh leaves were collected from a single individual and used for DNA extraction using the cetyltrimethyl ammonium bromide method [28]. DNA was quantified with a NanoDrop 2000c spectrophotometer (Thermo Fisher Scientific, Waltham, MA, USA) and was evaluated by electrophoresis in 1.0% agarose gel.

2.5. Sequencing library construction and high-throughput resequencing

After the quality of the DNA samples was checked, the samples were randomly cleaved into 200–500 bp fragments by sonication. Pair-end libraries with insert sizes of ~400 bp were constructed using the NEBNext Ultra DNA Library Prep Kit for Illumina (New England Biolabs, Ipswich, MA, USA) in accordance with the manufacturer's instructions and sequenced on an Illumina HiSeq 2500 platform (Illumina, San Diego, CA, USA) with the paired-end 2 × 150 bp mode.

2.6. SNP identification and genotyping

Raw reads from the HiSeq 2500 system were filtered to obtain high-quality reads. The clean reads were mapped to the reference genome. The program used to filter the raw reads was an in-house Perl script from the Biomarker Technology Co. Ltd. (Shunyi, Beijing, China). First, reads with adapter sequences and reads that contained >10% N were filtered, then low-quality reads with >50% of the bases with Q ≤ 10 were removed. To analyze the genomic variation, the clean reads from each sample were counted, and the coverage ratio and distribution of the reads in the TO1000 kale reference genome (https://plants.ensembl.org/Brassica_oleracea/Info/Index?db=core) [14] were calculated using the Burrows-Wheeler Aligner program [29]. The "CollectInsertSizeMetric" function in the Picard software toolkit (http://broadinstitute.github.io/picard/) was used to analyze the insert size. The GATK software toolkit was used to detect potential SNPs [30]. The Mark Duplicate tool was used to eliminate duplicated reads that may have resulted from PCR amplification. GATK was used for base recalibration, variant calling, and to strictly filter the SNPs to obtain the final SNP clusters

[31]. The default parameters were used for SNP calling and mapping for the parental lines and the F_1DH population. The SNPs identified between the parents were considered polymorphic in the subsequent bin calling. To guarantee the quality of the genetic map, the SNPs were filtered using the following criteria: 1) select markers that are homozygous and inconsistent in the parents; 2) the depth of parental markers is not less than 4×; and 3) remove markers that are not mapped on chromosomes. Gene genotyping of the progenies was based on the parental genotype. Only biallelic SNPs were retained in the final SNP data set. A chi-square test was conducted for bin calling; SNPs that significantly deviated from an extreme segregation distortion ($p < 0.001$) were excluded.

2.7. Genetic map construction and evaluation

The HighMap software (http://highmap.biomarker.com.cn/) was used to perform linkage grouping, marker ordering, genotyping error correction, and for mapping evaluation [32]. A modified sliding window method was used for bin calling [33]. The SNPs between the adjacent recombinant breakpoints were combined into a recombination bin. The bin markers were used to construct linkage groups with HighMap. The SMOOTH error correction strategy was performed based on the parental contribution of the genotypes [34], and the Kosambi mapping function was used to estimate the map distances. The ALLMAPS program was used to construct the chromosomes by comparing the constructed genetic map and the kale genome map [35]. Haplotype and heat maps were constructed using the "draw haplotype-map.pl" and "draw heatmap.pl" Perl scripts, respectively.

### 2.8. Relationship between the genetic and physical maps

The sequences of all bin markers used to construct the linkage map were aligned to the physical sequences of the TO1000 kale reference genome [14]. Collinearity between genetic and physical positions was determined by plotting genetic marker positions (in centimorgans) against their physical positions (in megabases). The BLAST program was used to confirm their physical positions in the genome. Spearman correlation coefficients were calculated to assess the collinearity between the genetic and physical maps [36].

3. Results

3.1. Microspore embryogenesis of ornamental kale F1 progeny

In this research, the F_1 progeny generated by crossing two DH lines, with strong differences in leaf shape and color (Figure 1), showed a high capability to produce embryos and develop into plants. The first cell division occurred on the fourth day after microspore culture. Cultured cells continued to divide to form, in turn, a cell group, original embryo, globular embryo, heart-shaped embryo, torpedo, and the cotyledon-stage embryo, which appeared after approximately 25 days. The differentiation and development of microspore cells were weakly synchronized in each Petri dish; even one month later, the embryoids in the same Petri dish were at different developmental stages (Figure 2a). Regenerated plants were formed from the episomal embryoids (Figure 2b) by different pathways. Some embryoids developed directly into regenerated plants (Figure 2c,d), whereas others formed secondary embryos through somatic embryogenesis of the hypocotyl, and then underwent further differentiation by secondary embryogenesis to generate adventitious buds that formed regenerated plants (Figure 2e,f).
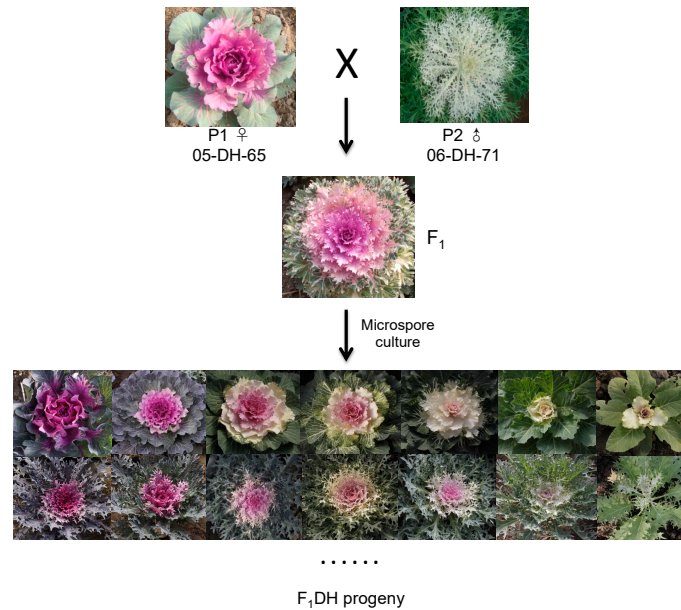
Figure 1. Phenotypes of the parental lines '05-DH-65' (female) and '06-DH-71' (male), the F1 progeny, and different morphotypes among the F$_1$DH progeny raised by microspore culture.
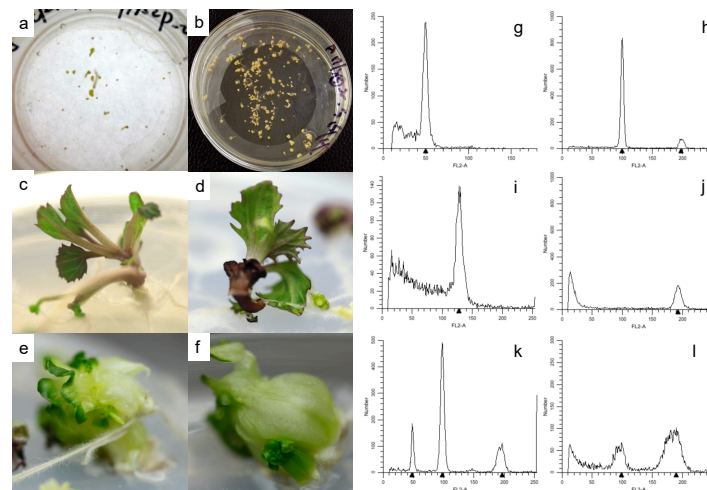
Figure 2. Embryoid formation and plant regeneration by microspore culture, and ploidy detection among regenerated plants derived from the ornamental kale F$_1$ progeny. Globular, heart, torpedo, and cotyledon-stage embryos in the same Petri dish (a). Episomal embryoids from microspore culture (b). Regenerated shoots directly generated from embryoids (c, d). Adventitious buds formed secondary embryos through somatic embryogenesis (e, f). Flow cytometry was used for ploidy detection among the regenerated plants (g–l).

3.2. Ploidy level and spontaneous chromosome doubling

In total, 1026 regenerated plants were obtained from microspore culture of the ornamental kale F$_1$ progeny. Their ploidy levels were determined by flow cytometry. The regenerated plants comprised 238 haploids (n, Figure 2g), 608 diploids (2n, Figure 2h), six triploids (3n, Figure 2i), 97 tetraploids (4n, Figure 2j), and 77 mixed-ploidy plantlets consisting of 55 "n+2n" plantlets (Figure 2k) and 22 "2n+4n" plantlets (Figure 2l). The relative proportions of diploids, haploids, triploids, tetraploids, "n+2n", and "2n+4n" plantlets were 23.21%, 59.26%, 0.58%, 9.45%, 5.36%, and 2.14%, respectively. The total percentage of spontaneous chromosome doubling (comprising diploids, polyploids, and mixed-ploidy plantlets) was 76.79%.

3.3. Construction of the F₁DH mapping population

From the 608 diploids, 300 genotypes that showed phenotypic variation were selected for construction of a F₁DH population of ornamental kale. The leaf color and leaf shape were investigated when plants were well colored. The colors of colored (inner) leaves of different F₁DH genotypes formed a continuum from white (no anthocyanin) to purple (more anthocyanin content), and the leaf shapes varied from entire (round) to deeply lobed (feather) (Figure 1). We self-pollinated these DH lines to obtain stable genetic progeny populations, and finally selected 150 inbred lines to form a mapping population for subsequent high-throughput whole-genome resequencing (WGR), genotyping, and high-density genetic map construction.

3.4. Whole-genome resequencing of the F₁DH mapping population and two parents

We performed WGR of the 150 F₁DH individuals and their two parents on an Illumina HiSeq sequencing platform. The raw reads were filtered, and low-quality sequences and redundant and unpaired reads were removed. In total, 145.12, 72.90, and 694.71 million clean reads were obtained for the parents '05-DH-65' and '06-DH-71' and the 150 F₁DH progenies, respectively, including 43.45, 21.83, and 208.09 Gb clean data with Q30 percentages of 91.45%, 85.59%, and 90.88%, respectively (Table 1). The respective GC contents were 37.59%, 36.74%, and 37.07%, and were normally distributed (Table 1 and Supplementary Figure S1). These results indicated sequencing data for the parents and the 150 F₁DH individuals were of good quality.

Table 1. Statistics for the whole-genome resequencing data

| Sample	Total Clean Reads	Total Clean Bases	Q30 Percentage (%)	GC Percentage (%)
05-DH-65	145,119,251	43,453,720,350	91.45	37.59
06-DH-71	72,896,717	21,830,253,446	85.59	36.74
F₁DH offspring	694,710,016	208,094,995,584	90.88	37.07
Total	912,725,984	273,378,969,380	90.84	37.07

The clean reads were aligned to the kale like rapid cycling reference genome (TO1000) [14] using the Burrows-Wheeler Aligner (BWA) software [29]. In total, 94.12%, and 92.59% of the total clean reads were mapped for the parents '05-DH-65' and '06-DH-71', respectively, and the average mapping rate for the F₁DH individuals was 93.55%. The percentages of mapped reads of the two parents and the average mapped reads of the progeny were all higher than 90%, and the properly mapped ratio (on average) was 80.85% and 78.45% for the two parents, and 81.41% for the progeny (Table 2). The average coverage depth was 64× and 34× for the two parents, and 2.6× for the progeny. There were more than 90% reads with coverage higher than 5× for the two parents, and more than 70% reads with average coverage higher than 1× for the progeny (Supplementary Table S1). The even distribution of reads coverage depth on the nine chromosomes indicated that the genome resequencing was sufficiently random (Supplementary Figure S2).

Table 2. Statistics for the mapping results

Sample	Clean Reads	Mapped (%)	Properly mapped (%)
05-DH-65	145,119,251	94.12	80.85
06-DH-71	72,896,717	92.59	78.45
F_1DH offspring (average)	9,199,180	93.55	81.41

3.5. SNP marker identification

Based on the genotype data for the two parents, SNP markers were identified. In total, 1,779,215 SNPs were detected between the two parental lines. Among these SNPs, 1,491,165 (83.8%) SNPs were distributed on the nine pseudochromosomes and 288,050 SNPs were on unanchored scaffolds. The SNPs were distributed throughout the nine pseudochromosomes with the highest number of SNPs occurring on chromosome C03 and the fewest on chromosome C06 (Figure 3). To ensure the quality of the map, the markers were filtered and screened (see Materials and Methods). After filtering, 1,007,400 SNPs were able to genotype the F_1DH population and were polymorphic. All of the SNP sites in the F_1DH population were integrated into a recombination bin unit, and 1,781 recombinant bins were used to construct the genetic map. To the best of our knowledge, the SNP markers that were mapped are more numerous than for previously reported genetic maps of *B. oleracea*, and form the first high-density genetic map of ornamental kale generated by WGR.
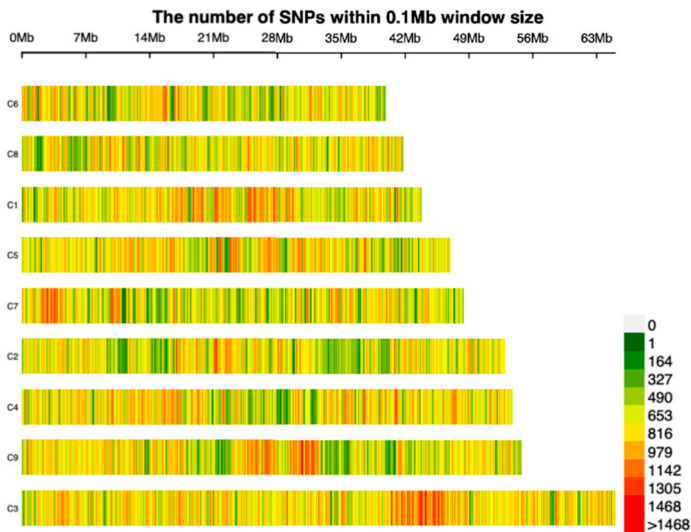

Figure 3. Heatmap of the distribution of SNPs across the nine chromosomes of the ornamental kale genome. A window size of 0.1 Mb was selected on the genome to count the number of SNPs. The color ranges from green to red indicating the density of SNPs from low to high.

3.6. Construction of the high-density bin genetic linkage map

To ensure the quality of the genetic map, we further polished and filtered the bins. The final high-density genetic map consisted of nine chromosomes and was based on 1,696 recombination bins containing 982,642 SNPs with a well-distributed distance (Figure 4). The length of the nine linkage groups ranged from 51.34 cM to 132.88 cM, and the total distance of the genetic linkage map was 775.81 cM, with an average distance of 0.46 cM between adjacent bin markers. The number of bin markers on the different chromosomes ranged from 126 to 272, and the average length of the bins was 0.295 Mb (Table 3).
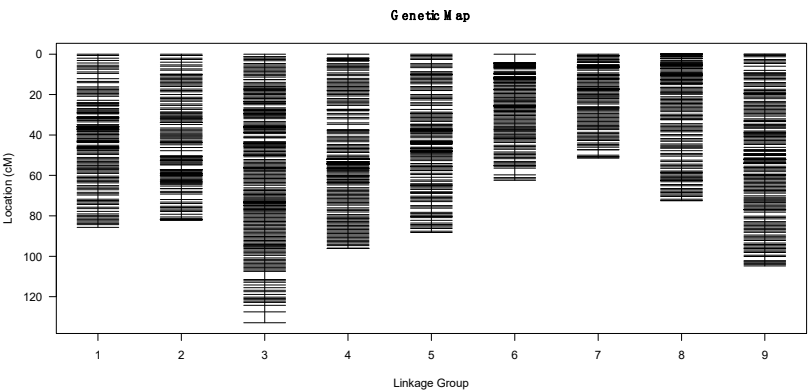
Figure 4. High-density genetic map of ornamental kale constructed from bin markers based on whole-genome resequencing.

Table 3 Distribution of genetic markers on the high-density genetic map

| LG ID	Total Bin Marker	Total Distance (cM)	Average Distance (cM)	Max Gap (cM)	Gaps<5cM (%)
LG1	156	85.73	0.55	2.67	100
LG2	164	81.89	0.5	2.73	98.36
LG3	272	132.88	0.49	5.35	99.73
LG4	221	96.06	0.43	2.67	98.89
LG5	179	88.25	0.49	2	100
LG6	170	62.4	0.37	3.34	100
LG7	126	51.34	0.41	2.67	100
LG8	180	72.45	0.4	2	100
LG9	228	104.81	0.46	0.02	100
Total	1696	775.81	0.46		99.66

In general, the bin markers were well-distributed on the genome and more than 99% of the intervals between adjacent markers were less than 5 cM. The largest gap, which occurred in LG3, was only 5.35 cM. LG3 was the largest linkage group, which covered 132.88 cM and included 272 bin markers, with an average distance of 0.49 cM between adjacent markers. The smallest LG was LG7, which covered 51.34 cM and included 126 bin markers with an average distance of 0.41 cM between adjacent markers (Table 3).

Haplotype and heat maps were used to assess the quality of the genetic map. The haplotype map, which counted the haplotype source in all linkage groups, reflected the double crossovers of the population and deletions, indicating recombination events and marker ordering errors. Haplotype maps were generated for each of the 150 lines in the F_1DH population as well as the parental lines. As illustrated in Figure 5, almost all of the recombination blocks were clearly defined.
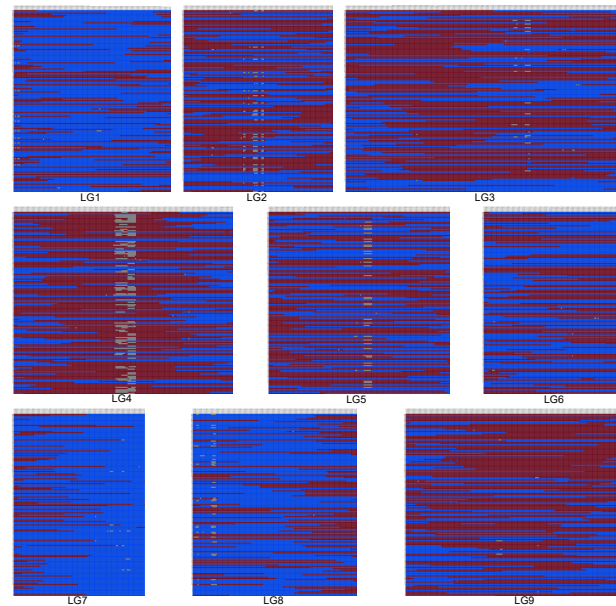
Figure 5. Haplotype map of the genetic map. Each row represents a marker, arranged in the order in the linkage map. Each column represents one sample. Red represents the female parent, blue represents the male parent, and gray indicates missing data. Where the color changes in a column is the location of a recombination event.

The genetic map is essentially a multi-loci recombination analysis. The recombination rate is lower if the marker is closer. The linkage relationship of the markers in the linkage group can be reflected by drawing heatmaps of marker relationships. Heatmaps of the nine linkage groups were generated separately based on pair-wise recombination values for the 1,696 recombination bin markers. The linkage relation heatmap presented in Figure 6 provides a visual representation of the relationship between recombination markers on a single chromosome, allowing for the identification of potential marker ordering errors. The heatmap utilizes colors to indicate the recombination rates between markers, with yellow indicating lower recombination rates and purple indicating higher rates. The heatmap analysis demonstrated that the construction of the genetic map was accurate, as the linkage groups were clearly defined and easy to visualize.
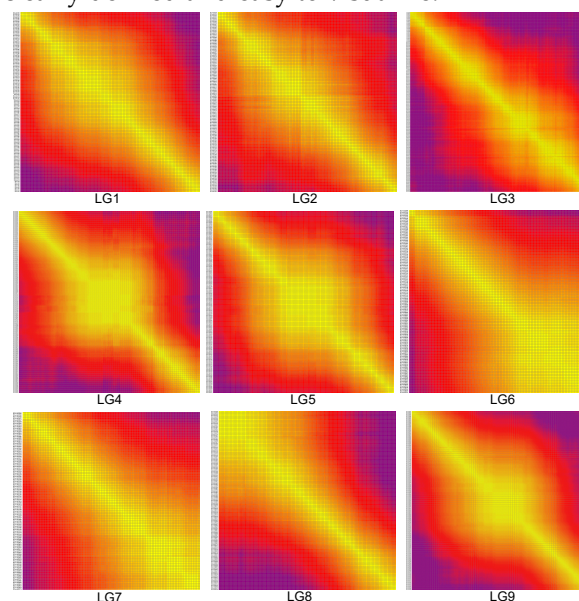

Figure 6. Heatmap of the genetic map. The x and y axes represent markers. The order of markers in a row and column are arranged according to their genetic distance. Each cell represents the

recombination rate of two markers. Yellow and purple indicate a low and high recombination rate, respectively.

3.7. Collinearity of the genetic and physical maps

To evaluate the collinearity between the genetic map and the kale reference genome (TO1000), all bin markers were mapped to the reference genome. As shown in Figure 7, the relationship between the genetic and physical maps were generally linear for all nine chromosomes, which indicated that the genetic map was consistent with the physical map. The Spearman correlation coefficients between the genetic and physical positions of the nine linkage groups were than more than 0.96 (Table 4). These results showed that these linkage groups show high levels of genetic collinearity with the physical map. The high collinearity suggested that the markers accurately cover the nine chromosomes and that they sufficiently cover the reference genome.
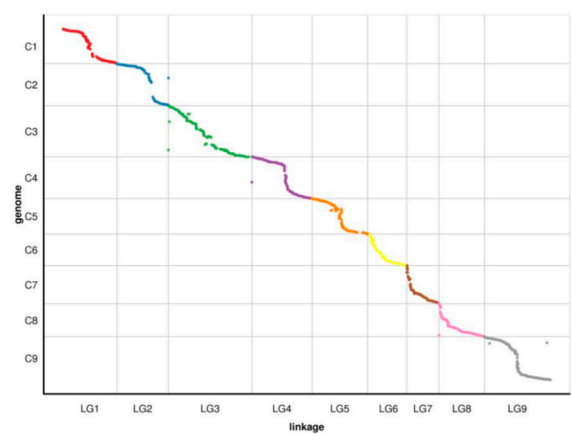
Figure 7. Relationship between genetic and physical positions for each chromosome. The x-axis represents the genetic distance on the genetic map, and the y-axis represents the physical position on the genome.

Table 4. Spearman correlation coefficients between the genetic and physical positions of each linkage group (LG).

LG ID	Spearman
LG1	0.9987
LG2	0.9926
LG3	0.9778
LG4	0.9816
LG5	0.9842
LG6	0.9976
LG7	0.9981
LG8	0.9641
LG9	0.9803

4. Discussion

4.1. Development of DH lines is an effective method for germplasm innovation of ornamental kale

The development of DH lines is an effective approach for germplasm innovation in plant breeding. The DH lines are genetically uniform and are true-breeding lines, and thus have the same genetic constitution as the original haploid plant from which they were derived. As a result, they are useful for development of new cultivars with desirable traits, such as disease resistance, improved quality, and higher yield. The DH lines can be generated by several methods, including microspore culture, anther culture, and ovule culture. These methods allow for the rapid production of

genetically uniform DH lines, which can be used to improve the efficiency of plant breeding programs [37]. By using DH lines, plant breeders can reduce the time required by traditional breeding methods, which can take several years. The DH lines can be used to evaluate and select desirable traits in a shorter period, allowing breeders to quickly develop new cultivars with improved characteristics.

Ornamental kale is a popular plant on account of its colorful foliage and unique appearance, but compared with other brassicaceous crops, limited germplasm resources are available for this plant. This can limit the potential for development of new and improved cultivars with desirable traits. Breeding ornamental kale can be challenging owing to its complex genetics and reproductive characteristics. This can make it difficult to develop new cultivars or to transfer desirable traits between varieties.

In this study, we developed a stable and efficient microspore culture system for ornamental kale to produce double haploids. The regenerated plants can be generated by different pathways from embryoids. The spontaneous chromosome-doubling rate was relatively high (>75%) during microspore regeneration, and the percentage of diploids was more than 20%. The F_1DH progenies exhibited rich phenotypic diversity, especially in leaf shape and color, and represented novel phenotypes that were not present in the parents. These results indicated that development of DH lines by microspore culture is an effective method for germplasm innovation of ornamental kale.

4.2. The F1DH population will play important roles in genetic analysis of ornamental kale

The F_1DH population is a type of mapping population that is commonly used in genetic map construction and QTL analysis. The F_1DH lines are homozygous at all loci and hence are ideal for genetic mapping and QTL analysis. This enables accurate and consistent mapping of genetic loci and precise identification of QTLs [4]. The F_1DH lines can be derived from F_1 hybrids that possess desirable traits, allowing for the efficient transfer of these traits to the mapping population. This is particularly useful in plant breeding, where desirable traits can be selected and incorporated into new varieties [38]. The homozygosity of F_1DH lines allows for the generation of high-density genetic maps and precise identification of QTLs. This is particularly useful in QTL analysis, where high-resolution mapping is required to identify the genetic loci responsible for complex traits [39].

Ornamental kale exhibits conspicuous variation in leaf color and shape, and is a suitable material to reveal the genes responsible. In this study, two ornamental kale DH lines differing markedly in leaf color and shape (one parent with red round leaves and the other parent with white feather leaves) were used as the parents and 300 diploid individuals were selected to construct the F_1DH population. The F_1DH progenies represented abundant genetic diversity in leaf color and shape, which exhibited a continuum from white to purple and from deeply lobed to entire. The F_1DH population provides an excellent genetic resource and will shorten the time required for breeding new varieties, and also lays the foundation for mapping the genes that control leaf color and shape in ornamental kale.

4.3. Features of the high-density genetic map of ornamental kale

A genetic map, which is a representation of the location of genes, genetic markers, and other genetic features on a chromosomes or genome, is important for finding loci of interest, fine mapping, gene identification, map-based cloning, comparative genomics, genome assembly, and MAS breeding [13,40]. The rapid development of NGS technologies, and the publication of the kale genome [14] and several other high-quality *B. oleracea* reference genomes [16–20], have made it possible to construct a high-density genetic map of ornamental kale using SNP markers and accurate genotyping. Several NGS-based genetic maps for *B. oleracea* have been published in recent years [21–23,41]. However, these genetic maps were all used simplified genome sequencing, such as genotyping by sequencing or SLAF markers. There are few reports of *B. oleracea* genetic maps based on WGR and no high-density genetic maps of ornamental kale have been reported previously. This sequence-based genotyping method is faster and more accurate than marker-based genotyping methods for constructing a genetic map and detecting recombination breakpoints [33,42]. The WGR mapping method significantly improves the efficiency of QTL mapping and marker development [43]. Based

on the WGR method, 982,642 SNP markers have been mapped on the genetic map. Many more SNP markers were mapped to the genetic map than in previous studies on other *B. oleracea* crops [21–23], which will provide valuable tools for future candidate-gene identification, map-based gene cloning, and MAS.

In this study, we sequenced both parental lines at a high coverage depth and sequenced the F_1DH offspring at a low coverage depth. A prominent feature of the WGR method for construction of genetic maps that was used in the present research is that it integrates SNP discovery, SNP validation, and genotyping [36,44,45]. We sequenced parental lines with more than 30-fold sequencing depth and each F_1DH progeny with an average 2.6-fold sequencing depth, which was sufficient to detect the recombination breakpoints. We obtained 43.45, 21.83, and 208.09 Gb of high-quality clean reads from the female parent, the male parent, and their progenies, respectively. By implementing rigorous analysis criteria, recombination intervals were removed, and bin markers were created using precise genotypic data to develop a high-density genetic map. The outcomes revealed that the WGR approach is an effective method for discovering markers and constructing linkage maps with high density. By using the WGR mapping technique, a vast number of genome-wide SNPs were identified that accurately represent the genomic and genetic diversity characteristics, and provided an abundance of polymorphisms for constructing the map [46,47].

In total, 1,696 recombination bin markers representing 982,642 SNP markers were mapped to nine linkage groups. The total length of the linkage map was 775.81 cM, with an average distance of 0.46 cM between adjacent bin markers. The collinearity of the genetic and physical maps was generally uniform for all nine chromosomes, which suggested that the markers accurately cover the chromosomes. Visual evaluation of the haplotype and heat maps of the genetic map suggested that the F_1DH population was suitable for genetic analysis. These results indicated that the high-density genetic map of ornamental kale contained a high marker density and was accurately constructed.

4.4. The application prospects of the high-density genetic map of ornamental kale

The high-density genetic map of ornamental kale can be applied in multiple areas of research and has excellent application prospects. It can be used in genome assembly and annotation to improve genome assembly and annotation accuracy by providing a scaffold for the assembly and identification of gaps or errors. It can be also used in comparative genomics to compare the genetic organization and evolution of different varieties of *B. oleracea* or other species of *Brassica*, which may provide insights into the mechanisms of speciation and genome rearrangement. Moreover, it can be used in QTL mapping to identify QTLs that control complex traits, such as leaf shape, leaf color, and disease resistance, in ornamental kale. It will also play important roles in the development of molecular breeding strategies for ornamental kale improvement, including MAS breeding, genomic selection, and gene editing. Overall, the high-density genetic map of ornamental kale is a powerful tool for genetic analysis and breeding, and its application is likely to continue to expand with continuing advances in genomic technologies.

5. Conclusions

In this study, a F_1DH population of ornamental kale was successfully constructed using microspore culture. The population was used to create a high-density genetic map, which contained a total of 1,696 recombination bin markers with an average marker density of 0.46 cM. These F_1DH progenies represent an excellent resource for germplasm innovation and the breeding of new varieties of ornamental kale. The high-density genetic map provides useful information for gene mapping and revealing the molecular genetic mechanism of important agronomic traits of ornamental kale. The present results provide valuable insights into the genetic basis of important traits in ornamental kale and provide a valuable resource for breeding and future research.

Supplementary Materials: The following supporting information can be downloaded at the website of this paper posted on Preprints.org, Figure S1: Base ratio distribution for the female parental line 05-DH-65, Figure S2: Coverage depth distribution on chromosomes of the female parental line. The *x*-axis represents the position

on each chromosome. The *y*-axis presents the $\log_2$ value of the coverage depth, Table S1: Coverage depth and coverage ratio of the sequencing data.

References

1. Li Y, Yu X: **Pollination with laser-irradiated pollens breaks cross-incompatibility between zicaitai (*Brassica campestris* var. *purpurea*) and ornamental kale (*Brassica oleracea* var. *acephala*) to produce hybrids with the aid of ovule culture**. *Scientia horticulturae* 2006, **108**(4):397-402.
2. Forster BP, Heberle-Bors E, Kasha KJ, Touraev A: **The resurgence of haploids in higher plants**. *Trends in plant science* 2007, **12**(8):368-375.
3. Dunwell JM: **Haploids in flowering plants: origins and exploitation**. *Plant biotechnology journal* 2010, **8**(4):377-424.
4. Forster BP, Thomas WT: **Doubled haploids in genetics and plant breeding**. *Plant breeding reviews* 2010, **25**:57-88.
5. Ferrie A, Caswell K: **Isolated microspore culture techniques and recent progress for haploid and doubled haploid plant production**. *Plant Cell, Tissue and Organ Culture (PCTOC)* 2011, **104**(3):301-309.
6. Zhang W, Fu Q, Dai X, Bao M: **The culture of isolated microspores of ornamental kale (*Brassica oleracea* var. *acephala*) and the importance of genotype to embryo regeneration**. *Scientia Horticulturae* 2008, **117**(1):69-72.
7. Yuan S, Liu Y, Fang Z, Yang L, Zhuang M, Zhang Y, Sun P: **Effect of combined cold pretreatment and heat shock on microspore cultures in broccoli**. *Plant Breeding* 2011, **130**(1):80-85.
8. Lv H, Wang Q, Yang L, Fang Z, Liu Y, Zhuang M, Zhang Y, Yang Y, Xie B, Wang X: **Breeding of cabbage (*Brassica oleracea* L. var. *capitata*) with fusarium wilt resistance based on microspore culture and marker-assisted selection**. *Euphytica* 2014, **200**(3):465-473.
9. Lou P, Zhao J, He H, Hanhart C, Pino Del Carpio D, Verkerk R, Custers J, Koornneef M, Bonnema G: **Quantitative trait loci for glucosinolate accumulation in *Brassica rapa* leaves**. *New phytologist* 2008, **179**(4):1017-1032.
10. Hirani AH, Li G: **Genetic Mapping, Quantitative Trait Analysis, and Gene Cloning in *Brassica oleracea***. In: *The Brassica oleracea Genome.* Springer; 2021: 7-22.
11. Gao M, Li G, Yang B, Qiu D, Farnham M, Quiros C: **High-density *Brassica oleracea* linkage map: identification of useful new linkages**. *Theoretical and applied genetics* 2007, **115**(2):277-287.
12. Wang W, Huang S, Liu Y, Fang Z, Yang L, Hua W, Yuan S, Liu S, Sun J, Zhuang M: **Construction and analysis of a high-density genetic linkage map in cabbage (*Brassica oleracea* L. var. *capitata*)**. *BMC genomics* 2012, **13**(1):1-12.
13. Jaganathan D, Bohra A, Thudi M, Varshney RK: **Fine mapping and gene cloning in the post-NGS era: advances and prospects**. *Theoretical and Applied Genetics* 2020, **133**(5):1791-1810.
14. Parkin IA, Koh C, Tang H, Robinson SJ, Kagale S, Clarke WE, Town CD, Nixon J, Krishnakumar V, Bidwell SL: **Transcriptome and methylome profiling reveals relics of genome dominance in the mesopolyploid *Brassica oleracea***. *Genome biology* 2014, **15**(6):1-18.
15. Liu S, Liu Y, Yang X, Tong C, Edwards D, Parkin IA, Zhao M, Ma J, Yu J, Huang S *et al*: **The *Brassica oleracea* genome reveals the asymmetrical evolution of polyploid genomes**. *Nat Commun* 2014, **5**:3930.

16. Cai X, Wu J, Liang J, Lin R, Zhang K, Cheng F, Wang X: **Improved *Brassica oleracea* JZS assembly reveals significant changing of LTR-RT dynamics in different morphotypes**. *Theoretical and Applied Genetics* 2020, **133**(11):3187-3199.

17. Lv H, Wang Y, Han F, Ji J, Fang Z, Zhuang M, Li Z, Zhang Y, Yang L: **A high-quality reference genome for cabbage obtained with SMRT reveals novel genomic features and evolutionary characteristics**. *Scientific reports* 2020, **10**(1):1-9.

18. Guo N, Wang S, Gao L, Liu Y, Wang X, Lai E, Duan M, Wang G, Li J, Yang M: **Genome sequencing sheds light on the contribution of structural variants to *Brassica oleracea* diversification**. *BMC biology* 2021, **19**(1):1-15.

19. Sun D, Wang C, Zhang X, Zhang W, Jiang H, Yao X, Liu L, Wen Z, Niu G, Shan X: **Draft genome sequence of cauliflower (*Brassica oleracea* L. var. *botrytis*) provides new insights into the C genome in *Brassica* species**. *Horticulture research* 2019, **6**.

20. Belser C, Istace B, Denis E, Dubarry M, Baurens F-C, Falentin C, Genete M, Berrabah W, Chèvre A-M, Delourme R: **Chromosome-scale assemblies of plant genomes using nanopore long reads and optical maps**. *Nature plants* 2018, **4**(11):879-887.

21. Lee J, Izzah NK, Choi B-S, Joh HJ, Lee S-C, Perumal S, Seo J, Ahn K, Jo EJ, Choi GJ: **Genotyping-by-sequencing map permits identification of clubroot resistance QTLs and revision of the reference genome assembly in cabbage (*Brassica oleracea* L.)**. *DNA research* 2016, **23**(1):29-41.

22. Zhao Z, Gu H, Sheng X, Yu H, Wang J, Huang L, Wang D: **Genome-wide single-nucleotide polymorphisms discovery and high-density genetic map construction in cauliflower using specific-locus amplified fragment sequencing**. *Frontiers in plant science* 2016, **7**:334.

23. Yu H, Wang J, Zhao Z, Sheng X, Shen Y, Branca F, Gu H: **Construction of a high-density genetic map and identification of loci related to hollow stem trait in broccoli (*Brassic oleracea* L. *italica*)**. *Frontiers in plant science* 2019, **10**:45.

24. Custers J, Cordewener J, Fiers M, Maassen B, van Lookeren Campagne M, Liu C: **Androgenesis in *Brassica*: a model system to study the initiation of plant embryogenesis**. *Current trends in the embryology of angiosperms* 2001:451-470.

25. Lv H-h, Wang Q-b, Yang L-m, Fang Z-y, Liu Y-m, Zhuang M, Zhang Y-y, Yang Y-h, Xie B-y, Wang X-w: **Breeding of cabbage (*Brassica oleracea* L. var. *capitata*) with fusarium wilt resistance based on microspore culture and marker-assisted selection**. *Euphytica* 2014, **200**:465-473.

26. Han S, Guo N, Zhang Y, Zong M, Wang G, Liu F: **Researches on the double haploid breeding of ornamental kale (*Brassica oleracea* var. *acephala*)**. *Journal of Agricultural Biotechnology* 2018, **26**(3):521-529.

27. Dpooležel J, Binarová P, Lcretti S: **Analysis of nuclear DNA content in plant cells by flow cytometry**. *Biologia plantarum* 1989, **31**(2):113-120.

28. Murray M, Thompson W: **Rapid isolation of high molecular weight plant DNA**. *Nucleic acids research* 1980, **8**(19):4321-4326.

29. Li H, Durbin R: **Fast and accurate short read alignment with Burrows–Wheeler transform**. *bioinformatics* 2009, **25**(14):1754-1760.

30. DePristo MA, Banks E, Poplin R, Garimella KV, Maguire JR, Hartl C, Philippakis AA, Del Angel G, Rivas MA, Hanna M: **A framework for variation discovery and genotyping using next-generation DNA sequencing data**. *Nature genetics* 2011, **43**(5):491-498.

31. Reumers J, De Rijk P, Zhao H, Liekens A, Smeets D, Cleary J, Van Loo P, Van Den Bossche M, Catthoor K, Sabbe B: **Optimized filtering reduces the error rate in detecting genomic variants by short-read sequencing**. *Nature biotechnology* 2012, **30**(1):61-68.

32. Liu D, Ma C, Hong W, Huang L, Liu M, Liu H, Zeng H, Deng D, Xin H, Song J: **Construction and analysis of high-density linkage map using high-throughput sequencing data**. *Plos one* 2014, **9**(6):e98855.

33. Huang X, Feng Q, Qian Q, Zhao Q, Wang L, Wang A, Guan J, Fan D, Weng Q, Huang T: **High-throughput genotyping by whole-genome resequencing**. *Genome research* 2009, **19**(6):1068-1076.

34. Van Os H, Stam P, Visser RG, van Eck HJ: **SMOOTH: a statistical method for successful removal of genotyping errors from high-density genetic linkage data**. *Theoretical and applied genetics* 2005, **112**:187-194.

35. Tang H, Zhang X, Miao C, Zhang J, Ming R, Schnable JC, Schnable PS, Lyons E, Lu J: **ALLMAPS: robust scaffold ordering based on multiple maps**. *Genome biology* 2015, **16**:1-15.

36. Li B, Lu X, Dou J, Aslam A, Gao L, Zhao S, He N, Liu W: **Construction of a high-density genetic map and mapping of fruit traits in watermelon (*Citrullus lanatus* L.) based on whole-genome resequencing**. *International journal of molecular sciences* 2018, **19**(10):3268.

37. Ferrie A, Caswell K: **Isolated microspore culture techniques and recent progress for haploid and doubled haploid plant production**. *Plant Cell, Tissue and Organ Culture (PCTOC)* 2011, **104**:301-309.

38. Demo BK: **Development of intervarietal substitution lines in *Brassica napus* L. using marker assisted selection and mapping of QTL for agronomically important traits**. Göttingen, Univ., Diss., 2007; 2007.

39.   Yadava SK, Ramchiary N: **Molecular Linkage Mapping in *Brassica juncea*: Founding the Basis for Marker-Assisted Selection**. In: *The Brassica juncea Genome*. Springer; 2022: 197-219.

40.   Boopathi NM, Boopathi NM: **Linkage map construction**. *Genetic Mapping and Marker Assisted Selection: Basics, Practice and Benefits* 2020:179-227.

41.   Yu H, Wang J, Sheng X, Zhao Z, Shen Y, Branca F, Gu H: **Construction of a high-density genetic map and identification of loci controlling purple sepal trait of flower head in *Brassica oleracea* L. *italica***. *BMC plant biology* 2019, **19**(1):1-8.

42.   M Perez-de-Castro A, Vilanova S, Cañizares J, Pascual L, M Blanca J, J Diez M, Prohens J, Picó B: **Application of genomic tools in plant breeding**. *Current genomics* 2012, **13**(3):179-195.

43.   Lim J-H, Yang H-J, Jung K-H, Yoo S-C, Paek N-C: **Quantitative trait locus mapping and candidate gene analysis for plant architecture traits using whole genome re-sequencing in rice**. *Molecules and cells* 2014, **37**(2):149.

44.   Xu X, Zeng L, Tao Y, Vuong T, Wan J, Boerma R, Noe J, Li Z, Finnerty S, Pathan SM: **Pinpointing genes underlying the quantitative trait loci for root-knot nematode resistance in palaeopolyploid soybean by whole genome resequencing**. *Proceedings of the National Academy of Sciences* 2013, **110**(33):13469-13474.

45.   Scheben A, Batley J, Edwards D: **Genotyping-by-sequencing approaches to characterize crop genomes: choosing the right tool for the right application**. *Plant biotechnology journal* 2017, **15**(2):149-161.

46.   Torkamaneh D, Boyle B, Belzile F: **Efficient genome-wide genotyping strategies and data integration in crop plants**. *Theoretical and Applied Genetics* 2018, **131**:499-511.

47.   Le Nguyen K, Grondin A, Courtois B, Gantet P: **Next-generation sequencing accelerates crop gene discovery**. *Trends in plant science* 2019, **24**(3):263-274.