***Brassica napus* L. dwarfing gene: Determining candidate intervals of dwarfing genes by BSA and SNP typing**

Luo Jing^{1,2}, Li Chao^{1,2*}, Zhang Ruimao², Chen Zhineng², Zhang Xianqiang³, Gao Zhihong², Lei Lei², Li Pan¹

1. School of Life Sciences, Guizhou Normal University, Guiyang, Guizhou, People's Republic of China, **2.** Rapeseed Institute, Guizhou Academy of Agriculture Sciences, Guiyang, Guizhou, People's Republic of China, **3.** Guizhou Police College, Guiyang, Guizhou, People's Republic of China.

*corresponding author

E-mail: gzlichao@126.com

Abstract

The plant height of rapeseed is one of the important factors that affects the production of rapeseed. If the plant height of rapeseed is too high, on the one hand, it will cause rapeseed to lodge and affect the yield, on the other hand, it will also affect the mechanized harvesting of rapeseed. In this research, the high-stalked line (YY50) and the dwarfed line (DW871) are crossed to obtain an F2 rapeseed population which was used to build pools, and then we used this to mine the main dwarfing genes. In the pools composed of tall and short stalks, we obtained 192.80Mb clean reads, which can be used for BSA (bulked segregant analysis). Preliminary positioning around the candidate section identified 23 SNP markers. Then 17 polymorphic SNP markers were obtained through polymorphism screening. Further we narrowed the candidate interval, and finally determined between 15.51-16.60Mb of ChrA10. Through identifying 231 genes from the above interval, it's predicted that the production of dwarf traits may be related to lignin synthesis and limited inflorescence. It provides a basis for further mapping and cloning of the dwarfing gene DW871.

Keyword: *Brassica napus* L. Plant height traits DW871 SNP typing
Bulked segregant analysis

1. Introduction

Rapeseed is an important oilseed economic crop. As China's main imported commodity, it is also the main exporting commodity of economies such as Canada and the European Union. In 2017/2018, the world's rapeseed production reached 74.92 MT^[1]. The main rapeseed producing areas are Canada, the European Union, and China. The planting area in 2017/2018 was 9.5 million hm², 6.6 million hm², and 6.8 million hm². By 2021, the global rapeseed harvest area is expected to reach 35.14 million hm²^[2]. The research of crop dwarf breeding is one of the main methods of the green

revolution. Rapeseed is another crop that dominates the green revolution besides rice and wheat [3]. By breeders' long-term efforts, production of rapeseed (*Brassica napus* L.) has been improved significantly. Correspondingly, lodging restricts the increase in yield and becomes one of the main factors in the reduction of rapeseed production. Therefore, it becomes a new goal for breeders that breeding dwarf varieties that are resistant to lodging and adapt to mechanized harvesting.

The earliest discovery and report of the dwarf mutants of rapeseed found by Williams and Hill [4], named Bn5-2 and Bn5-8. In China, Shi [5] gets dwarf mutant ds-1 and ds-2 by EMS (Ethylmethylsulfon, 0.25%) treated breeding material. Zhang [6] obtained the dwarf line DW871 through cross configuration breeding, which can be inherited stably. DW871 is a semi-dwarf variety which has definite inflorescence. According to different planting methods, the average plant height is between 70cm-139.1cm. It's more total leaf numbers, shorter internodes, thicker xylem, lower branch position, and more compact plant type compare to wild type control. By identifying the dwarfing gene of DW871, we can make better use of this excellent germplasm resource and provide help for subsequent breeding work.

Identifying molecular markers and possible candidate genes related to the dwarfing traits of rapeseed is an important means for us to make full use of germplasm resources to increase yield and mechanization rate [7]. When traits are controlled by a single gene, molecular marker location is our main means to find specific traits and assisted breeding. QTL (quantitative trait locus) is often used in association analysis related to traits. By associating the marker with the target trait, we can analyze the degree of correlation between the relevant data. Before we proceed to the next step, such as gene mapping, functional analysis and gene cloning, we need to find molecular markers that are closely related to the target traits [8].

At present, many results have been achieved in the research of dwarfing genes in rapeseed. These dwarfing genes affect plant height features of rapeseed through many ways such as hormone regulation, polyamines, plant cell wall synthesis, and transcriptional regulation [9-16]. By QTL mapping, Liu [5] mapped the dwarfing gene *ds-1* between SSR markers SR12156 and SCM07, with a distance of 2.8cm from SR12156 and 0.7cm from SCM07. Wang [17] used the dwarfing material M176 and found the dwarfing gene located between SSR markers Na10G06 and BrA09Y21 on the ninth chromosome. The distance from Na10G06 is 7cm and the distance from BrA09Y21 is 6.5cm. Liao [18] studied the mutant GA2ox6 and determined that the gene was located on *BnaA10g1070D* on chromosome A10.

The second-generation sequencing technology (next-generation sequencing, NGS) is based on the enzymatic cascade chemiluminescence reaction of four enzymes (DNA polymerase, sulfurylase, luciferase, adenosine triphosphatase) in the same PCR reaction system, called pyrosequencing too [19]. BSA (Bulked Segregant Analysis) based on the hybridization of parents with significant differences in traits. In the segregating population of offspring, a certain amount of phenotypic differences is selected and DNA is mixed to construct two pools. Because the selected object has only the target trait, only the region containing the target gene at the genome level is different between the two pools, and this genetic difference can be used as a molecular marker linked to

the target trait [20]. The emergence of NGS and BSA led to the development of NGS-based BSA analysis methods, such as QTL-seq, Mutmap, SHOREmap, etc. They are widely used in gene location and cloning [21-23]. By QTL-seq to perform high-depth sequencing of the DNA pool of extreme traits, and then by analyzing the Δ (SNP-index) and G' value, we can determine the candidate interval of the DW871 dwarf gene [24,25].

SNP (single nucleotide polymorphism) is a new type of molecular marker widely used in character identification of rapeseed. Most SNPs are biallelic, genetically stable, and have high reproducibility [26]. Compared with other types of molecular markers such as SSR, polymorphism information contained in SNPs is lower, but application of high-throughput sequencing technology makes it possible to identify a large number of SNPs in multiple genomes at the same time. And provide conditions for the development of markers for identification traits, genetic diversity and mapping [27]. In addition, the emergence of SNP-related technologies such as SNP chips and SNP arrays that can be used for genotyping have expanded the application range of SNP molecular markers [28,29]. Positional cloning is also called map-based cloning. It is a method of separating genes based on the strategy of forward genetic, based on the genetic and physical map of the phenotype of the mutant individual and the phenotype site [30]. For example, Malmberg [31] used low-coverage nanopore long-read sequencing technology to evaluate the SNP genotyping of DH(doubled haploid) rapeseed. Sun [32] used SNP molecular markers to analyze QTLs related to the weight and shape of rapeseed seeds.

DW871 is a new dwarf rapeseed strain with excellent shape. By backcrossing and crossing with other *Brassica napus* high-stalk materials to calculate the segregation ratio of progeny, we found that the plant height trait of DW871 is a dominant pair gene control, there is no cytoplasmic effect, and it is affected by modified genes, which is manifested as the genetic characteristics of quantitative traits [33]. In this research, QTL-seq and BSA were used to analyze DW871 to determine the candidate interval of its dwarf trait [34].

2. Material and method

2.1 plant material

In this experiment, the dwarf *Brassica napus* L. DW871 was used as the experimental material and named 312bulk1, and the high-stalk material YY50 was named 312bulk2 as the control material(Fig 1). The environment has a great influence on the plant height traits of *Brassica napus* L., and the materials used are all taken from the Guizhou Academy of Agricultural Sciences which lives in Huaxi District, Guiyang, Guizhou Province, 106°59' east longitude and 26°01' north latitude. This place belongs to a subtropical humid and temperate climate, with an altitude of 1150m, an average precipitation of 84.4mm, an average temperature of 13.70°C, and a sunshine duration of 100.90h.

The method of transplanting seedlings is adopted environment that row spacing is 40cm, the plot is 20m², and the density is 300,000 plants/hm². The fertility level before and after transplanting is consistent with the field pest management.

2.2 Extraction of whole genome DNA and construction of plant height gene pool

According to the traits of field plants, the whole genome of 30 plants of high-stalk material and 30 plants of short-stalk material were extracted. After mixing according to the same ratio, gene pools for dwarf population (312bulk1) and a gene mixing pool for tall population (312bulk2) were constructed.

By the improved CTAB (Hexadecyl trimethyl ammonium Bromide) to extract the whole genome DNA. After quality and concentration identification, experiment adjust the DNA concentration, and finally mix with equal volume to establish two extreme plant height trait gene pools, respectively dwarf and tall gene traits [35].

2.3 Construction of DNA library

Collect 30 leaves of high-stalk and short-stalked materials respectively to ensure that the leaves are intact and free of disease or damage. And quickly process plant materials at -80°C. And use the improved CTAB method to extract the whole genome DNA of the test material. Detect the quality of the whole genome DNA to ensure the next step. Build a PE library for each sample DNA. Then we randomly break the genomic DNA of each sample to 300-500bp fragments, connect the sequencing adapters, and prepare DNA clusters. Select the sequencing strategy PE150 on Illumina HiSeq X-ten (China, Illumina Company) to sequence the whole genome. In the end, a high-depth sequencing result of a mixed pool of single-strain DNA with extreme traits was obtained.

2.4 Compare clean data with reference sequence and detect variation

Obtaining the high-quality clean data helps to improve the quality of the data and the credibility of the follow-up results [36]. By performing data quality control on the library to obtain one clean data by Cutadapt (version 1.13) and Trimmomatic (version 0.36). Then compare the clean data to the reference genome (Darmor (v4.1)) by BWA (version 0.7.15-r1140), which is downloaded from www.genoscope.cns.fr/brassicanapus/data/ [37]. In order to obtain the BAM file for variant calling, we use the comparison result in SAM format to convert it through Samtools and Picard tools. Finally, detecting mutations by GATK, we get candidate SNPs and InDels [38].

2.5 Determine the candidate interval for dwarfing genes

In order to determine the candidate interval of dwarfing genes, we screen the genome based on Δ (SNP-index). And through the G' value, we can effectively eliminate the "high frequency" noise associated with sequence variation. Get narrower candidate section. The formula [25] for calculating the value of Δ (SNP-index) and G'

is (1)-(5):

$$\text{SNP} - \text{index}(\text{aa}) = \text{Maa}/(\text{Taa} + \text{Maa}) \quad (1)$$

$$\text{SNP} - \text{index}(\text{bb}) = \text{Mbb}/(\text{Tbb} + \text{Mbb}) \quad (2)$$

$$\Delta(\text{SNP} - \text{index}) = \text{SNP} - \text{index}(\text{aa}) \quad (3)$$

$$k_j = \frac{(1 - D_j^3)^3}{S_w} \quad (4)$$

$$G' = \sum_{j \in w} K_j G_j \quad (5)$$

In the above formula, M represents the plant height trait, and T represents the highest marker, and aa and bb represent the genotypes of 312bulk1 and 312bulk2, respectively. Maa and Mbb represent the concentrations of the aa population derived from M and the bb population of M, and Taa and Tbb represent the depth of aa and T bb derived from T, respectively. $\Delta(\text{SNP-index})$ indicates the degree of correlation between the marker and the trait. The closer $\Delta(\text{SNP-index})$ is to 1, the stronger the correlation between the marker and the trait. D_j is the standard distance, S_w is the sum of $(1 - D_j^3)^3$ of all SNPs in W, G_j represents the average number of G mutants on related SNPs, and G' is the weighted average of G_j . In the actual operation process, we use QTLseqr (v0.7.3) to process the above formula and set the size of the sliding window to 2Mb. In addition, when analyzing the two sample pools according to the reading depth to verify the significance level of the relevant SNP, it needs to meet $P < 0.01$.

In the process of determining the candidate interval, the conditions used are as follows: (1) The sequencing depth of the two parents is greater than 8; (2) The sequencing depth of the two pools is greater than 10; (3) The SNP-index value of the two pools cannot be greater than 0.8 or less than 0.2 at the same time; (4) Polymorphism between parents.

Based on the SNP detected by BSA, we use the principle of map-based cloning^[39]. We will manually select 23 SNPs under the condition that the target range is covered as uniformly as possible and specific amplification primers can be designed. Capture the target DNA fragment by designing primers to establish a PE library. Secondly, based on the sequencing data of each site, we analyze the genotype of each target site, find the Individual exchanges among them, and determine the effective Individual exchange. Finally, comprehensively analyze multiple effective exchange individual plants to determine the candidate interval of the target gene.

3. Conclusion

3.1 Sequencing data and statistics of whole genome comparison results

This experiment was carried out in Hubei GENOSEQ Technology Co., Ltd., using Illumina Hiseq X-ten to sequence the DNA of two samples. In order to ensure the

smooth progress of the follow-up work, remove the clean reads of low quality, containing joint boxes, and containing more N. A total of 192.80Mb sequencing data was obtained, 312bulk1 pool totaled 82.95Mb, and 312bulk2 pool totaled 109.85Mb. The average GC content of the two pools was 38.07%, and the average Q30 ratio was 92.71% (Table 1). The nucleobase quality of most sequencing sites in samples 312bulk1 and 312bulk2 are above Q30, and the nucleobase content is balanced, and the sequencing results are stable.

Table 1: Sample sequencing data

Sample	Reads	Bases	GC(%)	Q20(%)	Q30(%)
312bulk1	82,954,523	23,852,763,079	37.99	97.50	92.53
312bulk2	109,848,541	31,636,636,178	38.14	97.63	92.88
Mean	96,401,532	27,744,699,629	38.07	97.57	92.71
Sum	192,803,064	55,489,399,257	-	-	-

Process the sample DNA through the Illumina library to obtain PE reads (paired-end reads). Then the PE reads were compared with the *Brassica napus* L. reference genome (Darmor(v4.1)). The ratio of sample 312bulk1 to the reference genome sequencing data was 92.82%, the average coverage was 74.80%, and the average coverage depth was 17.14X; the comparison ratio of sample 312bulk2 was 93.34%, the average coverage was 73.88%, and the average coverage depth was 22.84X (Table 2). The contrast ratio of the sample is higher than 70%, and the coverage depth is greater than 10, which meets the requirements. The sequencing of the two samples has good randomness, and the sequencing data achieve uniform coverage of the reference genome.

Table 2: Statistics of comparison results of sequencing data

Sample	Total reads	uniquely mapped		Repeatedly mapped		Unmapped		Coverage(%)	Mean Depth
		Count	Rate	Count	Rate	Count	Rate		
312bulk1		52,352,45	63.11	24,645,94	29.71	5,956,12	7.18		
1	82,954,523	7	%	3	%	3	%	74.80	17.14
312bulk2		68,387,27	62.26	34,140,63	31.08	7,320,63	6.66		
2	109,848,54	3	%	0	%	8	%	72.95	22.84
		60,369,86	62.69	29,393,28	30.40	6,638,38	6.92		
Mean	96,401,532	5	%	7	%	1	%	73.88	19.99

3.2 Screening of mutation sites by SNPs and InDels

Using GATK to complete the mutation detection of the sample DNA, we tested a total of 645.40Mb of data, and screened out 1,189,864 usable SNPs and 217,989 usable InDels. The labeling density of SNPs and InDels on A10 was 2847.13 and 682.83, and the labeling density was the highest in each chromosome (Table 3). Through the distribution map of available SNP/InDel sites on each chromosome, we can intuitively

see that the markers are concentrated on A01 and A10 (Fig 2). It indicates that the markers associated with genes related to dwarfing traits of *Brassica napus* L. mainly exist on chromosome A10.

Table 3: Samples can be chromosome-divided statistics by SNP and InDel

Chr	Length	SNPs		INDELs	
		Count	Density (Mb ⁻¹)	Count	Density (Mb ⁻¹)
A01	23,267,856	59,509	2557.56	14,075	604.91
A02	24,793,737	39,798	1605.16	8,346	336.62
A03	29,767,490	42,583	1430.52	10,389	349.00
A04	19,151,660	32,691	1706.95	7,275	379.86
A05	23,067,598	36,457	1580.44	7,209	312.52
A06	24,396,386	34,581	1417.46	6,266	256.84
A07	24,006,521	34,665	1443.98	7,478	311.50
A08	18,961,941	30,037	1584.07	6,285	331.45
A09	33,865,340	51,974	1534.73	10,964	323.75
A10	17,398,227	49,535	2847.13	11,880	682.83
C01	38,829,317	50,388	1297.68	9,425	242.73
C02	46,221,804	92,105	1992.67	14,362	310.72
C03	60,573,394	110,836	1829.78	19,723	325.61
C04	48,930,237	94,559	1932.53	14,715	300.73
C05	43,185,227	87,025	2015.16	13,040	301.96
C06	37,225,952	72,051	1935.50	11,709	314.54
C07	44,770,477	85,837	1917.27	13,645	304.78
C08	38,477,087	81,256	2111.80	13,556	352.31
C09	48,508,220	103,959	2143.12	17,647	363.79
Whole	645,398,471	1,189,846	1835.97	217,989	352.97

3.3 Preliminary positioning of candidate intervals

We determined the interval of $\Delta(\text{SNP-index}) > 4$ (Fig 3) and G' value > 2 (Fig 4), and determined the common interval between the two after comprehensive analysis. We located the candidate interval related to DW871 plant height traits between BnaA10g21330D and BnaA10g26440D on ChrA10. Starting at 14.70Mb and ending at 16.86Mb, the length of the candidate interval is 2.16Mb. In this range, we screened a total of 14,340 variants in the 312bulk1 and 312bulk2 genomes, including 512 genes, and 290 genes have large effects.

3.4 Fine positioning of candidate intervals

The candidate intervals are between BnaA10g21330D and BnaA10g26440D on ChrA10. According to the method in 2.5, we sequenced a total of 500 samples and obtained a total of 13.47Mb clean PE reads. The GC value is 42.42, the Q30 value is 98.03, and the average coverage depth of the site is 2453.76X. Meet the requirements.

In order to further narrow the candidate interval, we need to determine the individual in the sample that has exchanged genotypes and phenotypes. Screening polymorphisms between dwarf and tall stalks, let us finally obtain 17 pairs of primers (Table 4, middle box). We analyzed the phenotypic composition of the samples and identified 13 effective exchange individual plants out of 500 samples.

Table 4: Statistics of genotype composition of target locus

Locus ID	Genotype frequency
V001172	C/T:0.6960 C/C:0.2920 ./.:0.0120
V005863	G/T:0.7040 T/T:0.2940 ./.:0.0020
V008377	C/G:0.5200 G/G:0.2880 C/C:0.1820 ./.:0.0100
V008871	A/G:0.5200 G/G:0.2880 A/A:0.1820 ./.:0.0100
V009975	A/C:0.5260 C/C:0.2860 A/A:0.1820 ./.:0.0060
V011639	T/T:0.9900 ./.:0.0100
V014371	C/T:0.5300 T/T:0.2800 C/C:0.1900
V016167	./.:1.0000
V017284	A/G:0.5320 G/G:0.2760 A/A:0.1860 ./.:0.0060
V018247	A/G:0.5340 G/G:0.2840 A/A:0.1740 ./.:0.0080
V019455	./.:0.5340 C/T:0.2180 T/T:0.1260 C/C:0.1220
V020674	A/G:0.5260 G/G:0.2920 A/A:0.1720 ./.:0.0100
V020897	A/C:0.5320 C/C:0.2760 A/A:0.1820 ./.:0.0100
V021920	A/G:0.5660 A/A:0.2740 G/G:0.1500 ./.:0.0100
V023557	C/T:0.5400 T/T:0.2720 C/C:0.1860 ./.:0.0020
V024819	G/T:0.5340 G/G:0.2820 T/T:0.1780 ./.:0.0060
V026167	C/T:0.5320 C/C:0.2880 T/T:0.1720 ./.:0.0080
V028408	G/T:0.5360 T/T:0.2820 G/G:0.1740 ./.:0.0080
V029327	C/T:0.5280 C/C:0.2900 T/T:0.1700 ./.:0.0120
V029614	A/G:0.5360 A/A:0.2940 G/G:0.1700
V029643	C/T:0.5280 T/T:0.2860 C/C:0.1740 ./.:0.0120
V030469	A/A:0.9700 ./.:0.0100 C/C:0.0100 A/C:0.0100
V031219	C/T:0.9820 C/C:0.0160 T/T:0.0020

We verified the individual exchanges obtained after screening through these molecular markers, and found that the individual exchanges with high stalk phenotype decreased by 1, at the marker V008377, decreased 1, V008871 decreased 1, V009975 decreased 2, V014371 decreased 3, V017284 decreased 4; on the other side of the interval, the individual exchanges with the tall stalk phenotype are 1 at V029643 and V029614. This result shows that the target gene is between the markers V017284 and V029614. Comprehensive analysis of multiple individual exchange, we can determine the target gene between V018247 and V029614. Start at 15.51Mb and end at 16.60Mb. Its physical distance is 1.09Mb.

Between V018247 and V029614, we identified 231 genes. Through blasts in NCBI, 77 of which are genes, we are currently unable to determine their specific functions. In addition, we identified 24 genes related to transcription, 12 genes related to resistance, 42 genes involved in protein kinase recombination, 43 genes involved in intracellular

metabolism, and 15 genes involved in cell growth and the apoptosis of process, and 5 genes related to protein modification. Some genes are related to Lignin anabolic pathway, petal growth and flowering factors.

4. Summary

The final purpose of a series of work by breeders is to obtain new varieties of rapeseed with high yield and mechanized harvesting. The continuous increase in the yield of rapeseed led to the appearance of lodging, which in turn caused the yield of rape to drop significantly, reducing the economic benefits of rape and the enthusiasm of farmers to plant. Therefore, the dwarf breeding of *Brassica napus* L. has become an important means to provide the mechanization rate of rapeseed production and increase farmers' income. BSA-seq is a new generation method that relies on NGS to determine the candidate interval of the target gene. In this research, two pools with extreme differences in traits were constructed, named 312bulk1 and 312bulk2. The whole genome sequencing data of these two pools is all obtained by NGS sequencing after mixing multiple samples in equal proportions. After obtaining the sequencing data, by analyzing the data and calculating its SNP-index, Δ (SNP-index) and G' value, we found the candidate interval of the DW871 dwarf gene on ChrA10 chromosome, 14.70Mb~16.86Mb. The length of the candidate interval is 2.16Mb, including 512 genes, including 290 genes with large effects.

Because the current candidate interval is too large, it cannot meet the needs of subsequent gene mapping and breeding. Therefore, we need to further narrow the target gene candidate interval. SNP markers are a new generation of molecular markers and have been widely used. SNP genotyping technology is a gene mapping technology developed based on SNP molecular markers. Through map-based cloning and SNP genotyping technology, we can further narrow the candidate range of target genes. Through manual selection, we selected 23 SNP markers among the SNPs obtained after BSA, and finally obtained 17 polymorphic markers through polymorphism screening. Then through the comparison and analysis of the marker and the individual exchanges, we finally determined the candidate interval of the target gene, 15.51Mb~16.60Mb, and its physical distance was 1.09Mb.

Between V018247 and V029614, we identified 231 genes. Among 77 genes, we are currently unable to determine their specific functions. In addition, we identified 24 genes related to transcription, 12 genes related to resistance, 42 genes involved in protein kinase recombination, 43 genes involved in intracellular metabolism, and 15 genes involved in cell growth and the apoptosis of process, and 5 genes related to protein modification. Some genes are involved in lignin anabolic pathway, petal growth and flowering regulation. Therefore, we conclude that the dwarfing trait may be related to the synthesis of lignin. In addition, because DW871 is a definite inflorescence, we speculate that the formation of dwarfing traits is related to the formation of definite inflorescence. This research provides a reference for further determining the target range of dwarfing genes and verifying the formation mechanism of related traits.

However, to determine the formation mechanism of dwarf traits and the cloning of dwarf genes still needs more in-depth work.

Thanks:

This research is co-sponsored by Guizhou science and technology key project(Qian Ke He Basis [2017]1415), Guizhou outstanding young science and technology talent project(Qian Ke He Platform Talent [2017] 5642), Guizhou support plan project(Qian Ke He bracing [2017] 2568), Guizhou RenCai JiDi(RCJD2018-14), Guizhou provincial engineering research center for plant conservation technology application, and Construction of laboratory platform of agricultural science and technology innovation center in karst mountains(Qian Ke Zhong Yin Di [2018]4001). In addition, special thanks to Wen Dezhi of Guizhou Normal University for his selfless help in the translation and writing of this article.

References:

- [1] Unite States Department of Agriculture (USDA), Oilseeds: World Markets and Trade in Sep 2019(R), 2019.
- [2] Unite States Department of Agriculture (USDA), Oilseeds: World Markets and Trade in Jul 2020(R), 2019.
- [3] Duan Q.Y, Effects of Planting Density and Nitrogen Fertilizer on Growth, Yield and Nitrogen Utilization of Dwarf Rape[D], Thesis for M.S., Sichuan Agricultural University, 2017, Supervisor: Wu Y.C.
- [4] Williams J.K, Kubelik A.R, Livak K.J, Rafalski J.A, Tingey S.V, DNA polymorphisms amplified by arbitrary primers are useful as genetic markers[J], Nucleic Acids Res, 1990.
- [5] Liu Chao, Cloning and functional analysis of dwarf gene Bnrga-ds in rapeseed (*Brassica napus* L.)[D], Dissertation for Ph.D., Huazhong Agricultural University, 2010, Supervisor, Wu J.S., and Liu K.D.
- [6] Zhang R.M, Li Chao, Chen D.L, Xiang Yang, Breeding of short stem, erect plant type DW871 in *brassica napus* L.[J], Seed, 2019,38(02):116-120+123.
- [7] Adrian V. Dalca, Michael Brudno, Genome variation discovery with high-throughput sequencing data[J]. Briefings in bioinformatics, 2010, 11(1): 3-14.
- [8] Liu L.Z, Cunmin Qu, Benjamin Wittkop, Bin Yi, Yang Xiao, He Y.J, Rod J. Snowdon, Jiana Li, A high-density SNP map for accurate mapping of seed fibre QTL in *Brassica napus* L.[J], Plos One, 2013, 8:c83052.
- [9] Amber L.H, Tohru A, Camille M.S, The roles of the GA receptors GID1a, GID1b, and GID1c in sly1 -independent GA signaling[J], Plant Signaling and Behavior, 2014, 9: e28030.
- [10] Muangprom A, Osborn T.C, Characterization of a dwarf gene in Brassica rapa, including the identification of a candidate gene[J], Theor. Appl. Genetics, 2004, 7(7): 1378-1384.
- [11] Chu Z.Q, Li Li, Song Li, Xue H.W, Advance on brassinosteroid biosynthesis and function[J], ZhiwuTongbao (Chinese Bulletin of Botany), 2006, 23(5): 543-555
- [12] Cheng Y.F, Dai X.H, Zhao Y.D, Auxin biosynthesis by the YUCCA flavin monooxygenases controls the formation of floral organs and vascular tissues in Arabidopsis[J],

Genes Dev, 2006, 20(13): 1790-1799

- [13] Kakehi J, Kuwashiro Y, Niitsu M, Takahashi T, Thermospermine is required for stem elongation in *Arabidopsis thaliana*[J], *Plant Cell Physiol*, 2008, 49(9): 1342-1349
- [14] Cheng Xi, Hao H.Q, Peng Li, Recent progresses on cellulose synthesis in cell wall of plants[J], *Redai Yaredai Zhiwu Xuebao (Journal of Tropical and Subtropical Botany)*, 2011, 19(3): 283-290
- [15] Roudier F, Fernandez A.G, Fujita M, Himmelspach R, Bomer G.H, Schindelman G, Song S, Baskin T.I, Dupree P, Wasteneys G.O, and Benfey P.N, COBRA, an *Arabidopsis* extracellular glycosyl-phosphatidyl inositol anchored protein, specifically controls highly anisotropic expansion through its involvement in cellulose microfibril orientation[J], *Plant Cell*, 2005, 17(6): 1749-1763
- [16] Heish T.H, Lee J.T, Yang P.T, Chui L.H, Charng Y.Y, Wang Y.C, and Chan M.T, Heterology expression of the *Arabidopsis* Crepeat/dehydration response element binding factor 1 gene confers elevated tolerance to chilling and oxidative stresses in transgenic tomato[J], *Plant Physiol*, 2002, 129(3): 1086-1094
- [17] Wang Yu, Genetic segregation analysis and QTL Mapping of the plant height trait in *Brassica napus* L.[D], Thesis for M.S, Nanjing Agricultural University, 2012, Supervisor: Guan R.Z.
- [18] Liao X.Y, Cloning GA2ox family genes from *Brassica napus* L. and study on *BNGA2ox6* gene function in *Arabidopsis thaliana*[D], Thesis for M.S., Hunan University, 2015, Supervisor, Liu X.M.
- [19] Gong Min, Shi-Hui Foo, Lifang Lin, Edison T. Liu, Baback Gharizadeh, Sanket Goel, Pyrosequencing enhancement for better detection limit and sequencing homopolymers[J], *Biochemical and Biophysical Research Communications*, 2010, 401(2010): 117-123.
- [20] R. W. Michelmore, I. Paran, And R. V. Kesseli, Identification of markers linked to disease-resistance genes by bulked segregant analysis: a rapid method to detect markers in specific genomic regions by using segregating populations[J], *Proceedings of the National Academy of Sciences of the United States of America*, 1991, 88: 9828-9832.
- [21] Tudor Eleri H, Jones D Marc, He Z.S, Bancroft Ian, Trick Martin, Wells Rachel, Irwin Judith A, Dean Caroline, QTL-seq identifies BnaFT.A02 and BnaFLC.A02 as candidates for variation in vernalization requirement and response in winter oilseed rape (*Brassica napus*)[J], *Plant Biotechnology Journal*, 2020, (2020)1-16.
- [22] Takagi Hiroki, Tamiru Muluneh, Abe Akira, Yoshida Kentaro, Uemura Aiko, Yaegashi Hiroki, Obara Tsutomu, Oikawa Kaori, Utsushi Hiroe, Kanzaki Eiko, Mitsuoka Chikako, Natsume Satoshi, Kosugi Shunichi, Kanzaki Hiroyuki, Matsumura Hideo, Urasaki Naoya, Kamoun Sophien, Terauchi Ryohei., MutMap accelerates breeding of a salt-tolerant rice cultivar[J]. *Nature biotechnology*, 2015, 33(5).
- [23] Schneeberger Korbinian, Ossowski Stephan, Lanz Christa, Juul Trine, Petersen Annabeth Høgh, Nielsen Kåre Lehmann, Jørgensen Jan-Elo, Weigel Detlef, Andersen Stig Uggerhø, SHOREmap: simultaneous mapping and mutation identification by deep sequencing[J]. *Nature methods*, 2009, 6(8).
- [24] Takagi Hiroki, Abe Akira, Yoshida Kentaro, Kosugi Shunichi, Natsume Satoshi, Mitsuoka Chikako, Uemura Aiko, Utsushi Hiroe, Tamiru Muluneh, Takuno Shohei, Innan Hideki, Cano Liliana M, Kamoun Sophien, Terauchi Ryohei, QTL-seq: rapid mapping of quantitative trait

- loci in rice by whole genome resequencing of DNA from two bulked populations[J]. *Plant J*, 2013, 74(1):174-83.
- [25] Magwene PM, Willis JH, Kelly JK, The Statistics of Bulk Segregant Analysis Using Next Generation Sequencing[J]. *PLOS Computational Biology*, 2011, 7(11): e1002255
- [26] Mansfeld B.N, Grumet R, QTLseqr: An R package for bulk segregant analysis with next-generation sequencing[J]. *The Plant Genome*, 2018.
- [27] Durstewitz G, Polley A, Plieske J, Luerssen H, Graner E M, Wieseke R, Ganal M W., SNP discovery by amplicon sequencing and multiplex SNP genotyping in the allopolyploid species *Brassica napus*[J], *Genome*, 2010, 53(11):948-956.
- [28] McKenna Aaron, Hanna Matthew, Banks Eric, Sivachenko Andrey, Cibulskis Kristian, Kernysky Andrew, Garimella Kiran, Altshuler David, Gabriel Stacey, Daly Mark, DePristo Mark A, The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data[J]. *Genome research*, 2010, 20(9): 1297-1303.
- [29] Wu Peng, Juan Liu, Tian J.C, Research Progress of Single Nucleotide Polymorphism (SNP) Molecular Markers in Wheat Genetic Breeding[J], *Journal of Agriculture*, 2019,9(1):54-58.
- [30] Tanksley SD, Ganal MW, Martin GB, Chromosome landing: a paradigm for map-based gene cloning in plants with large genomes[J], *Trends Genet.* 1995, 11(2):63-68.
- [31] M. M. Malmberg, Assessment of low-coverage nanopore long read sequencing for sNP genotyping in doubled haploid canola (*Brassica napus* L.) [J], *Science Reports*, 2019, 9(1).
- [32] Sun L.J, Wang X.D, Yu K.J, Li W.J, Peng Qi, Chen Feng, Zhang Wei, Fu S.X, Xiong D.Q, Chu Pu, Guan R.Z, Zhang J.F, Mapping of QTLs controlling seed weight and seed-shape traits in *Brassica napus* L. using a high-density SNP map[J], *euphytica*, 2018, 214:228.
- [33] Zhang R.M, Li Chao, Chen D.L, Xiang Yang, Genetic analysis of plant height in new dwarf line DW871 of *Brassica napus* L[J], *Chinese Journal of Oil Crop Sciences*, 2019, 41(1): 001-009
- [34] Das Shouvik, Deploying QTL-seq for rapid delineation of a potential candidate gene underlying major trait-associated QTL in chickpea, 2015.
- [35] Zhang X.Y, Wang Ling, Shou L.L, A Rapid Modified CTAB Method of Extracting Genomic DNA from Wheat Leaf[J], *Chinese Agricultural Science Bulletin*, 2012, 28(36): 46-49.
- [36] Marvin N. Wright, Damian Gola, Andreas Ziegler, Preprocessing and Quality Control for Whole-Genome Sequences from the Illumina HiSeq X Platform[J]. *Methods in molecular biology* (Clifton, N.J.), 2017, 1666.
- [37] Chen Y.P, Schmidt Bertil, And Maskell Douglas L, A hybrid short read mapping accelerator[J], *BMC bioinformatics*, 2013, 14.
- [38] Anonymous, Genomics; New genomics findings from Broad Institute of Harvard and MIT described[J], *Science Letter*, 2010.
- [39] Coulson A, Sulston J, Brenner S, Karn J, 1986, Toward a physical map of the genome of the nematode *Caenorhabditis elegans*. [J]. *Proceedings of the National Academy of Sciences of the United States of America*, 1986, 83(20).

Fig 1: Comparison chart of DW871 and yy50

Fig 2: The x-axis is the chromosome position, and the y-axis is the

Fig 3: $\Delta(\text{SNP-index})$

Fig 4: G' value