Whole Genome Resequencing Enables the Exploration of "Mysterious Factors" Regulating Crop Height

Drawing a high-density genetic map based on MGI's DNBSEQ platform to determine a major QTL locus controlling height in rapeseed.

Researchers from Oil Crops Research Institute of the Chinese Academy of Agricultural Sciences and National Center of Rapeseed Improvement, Huazhong Agricultural University used whole genome sequencing (WGS) to conduct quantitative trait locus sequencing (QTL-seq) and genetic analysis for *Brassica napus* and achieved new progress in the study of *Brassica napus* height. Sequencing platform of MGI Tech Co., Ltd. (MGI) has contributed to a portion of the sequencing work.

The research entitled "Mapping of a major QTL controlling plant height using a high-density genetic map and QTL-seq methods based on whole-genome sequencing in Brassica napus" has been published in the internationally renowned journal *Genes Genomes Genetics* in 2021[1].

Recommended application: Molecular breeding

Recommended model: DNBSEQ-G400RS

- **Data output is efficient and high-quality**

  DNBSEQ sequencing technology has significant features such as high accuracy, low repeat sequence rate, and low index hopping rate, etc.

- **Offer a complete product combination for experimental procedures**

  Based on independently developed automatic solutions and analysis software, MGI provides a complete set of product combination from sample to result output.

Background

Plant height is a key factor influencing crop yield[1]. A major successful case of plant breeding was the introduction of the dwarf loci into crops in the 1960s. Employing dwarf loci in crops enhanced the lodging resistance, facilitated reaping, and significantly increased yield. This progress is named the "Green Revolution"[2]. Rapeseed (*Brassica napus*) is one of the most important oil crops globally, and its relative trait "tall and dwarf" not only affects the yield of rapeseed, but also affects several important agronomic traits such as resistance to overturning and ease of harvesting. Therefore, studying the height of *Brassica napus* has significant implications for agronomic practices[3].

In recent years, scientists have identified 183 height-related quantitative trait loci (QTLs) in *Brassica napus*, which can explain 3%–70% of phenotypic variation in plant height[1, 4]. Plant height is a quantitative trait jointly controlled by environment and numerous genes. As traditional method for QTL mapping is time-consuming and labor-intensive, a novel approach to study *Brassica napus* height is highly demanding[5].

With the development of whole genome sequencing (WGS) technology, and the announcement of *Brassica napus* reference genome, as well as the density elevation of molecular markers in linkage maps, the positioning of QTL is becoming more accurate[6]. Genome-wide association study (GWAS) and QTL-seq have been applied to crop genetic breeding. Abundant single nucleotide polymorphisms (SNPs) from the whole genome can overcome the limitations of traditional molecular markers, including high workloads, low accuracy, and so on. Therefore, a sequencing platform with high accuracy and sequencing depth becomes a powerful tool for agricultural genetic breeding[7].

Study description

In this study, a QTL controlling the height of *Brassica napus* was identified using QTL-seq and whole-genome resequencing. By integrating genetic variations and gene expression changes information, candidate genes regulating *Brassica napus* height in the QTL interval were identified. This provides a significant resource for cultivar elevation and genetic breeding of *Brassica napus*. The MGI sequencing platform undertook the whole genome resequencing and RNA-seq related work.

Materials and Methods

A. Materials

Zhongshuang 11-HP (ZS11-HP), a tall mutant, and semi-dwarf mutant (sdw-e), a dwarf mutant, were conducted parental cross to construct F_2 segregating population. We planted 200 F_2 population seedlings as genetic analysis strains, and constructed two extreme plant height bulks by selecting the 20 tallest and 20 shortest plants (Figure 1).

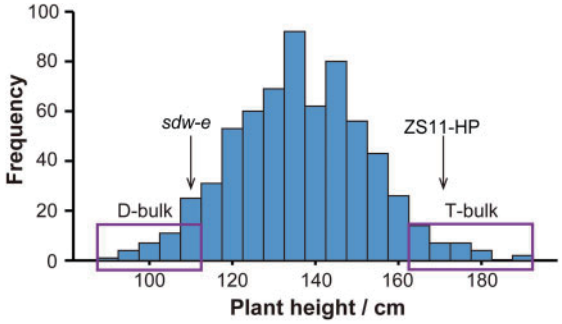

Figure 1. Phenotypic distribution of the F_2 segregating population and selection criteria for the extreme bulks.

B. Library preparation and sequencing

Fresh leaves of the two parents and F_2 population in the seedling stage were collected to conduct high-quality DNA extraction and library construction. Shoot apical meristems (SAMs) of the two parents in the bolting stage with two biological replicates were harvested to carry out RNA extraction and RNA library preparation. Vendor A platform was used to perform the QTL-seq for the two parents and two extreme bulks selected from the F_2 population, whereas the MGI platform was utilized to perform whole genome resequencing for the 200 individuals in the F_2 population and RNA-seq for SAMs of the two parents.

MGIEasy Universal DNA Library Prep Set (16 RXN) and MGIEasy RNA Library Prep Set (16 RXN) are recommended for library preparation. MGI automated platform is recommended for performing DNA/ RNA extraction and library construction, which can effectively improve trial efficiency. For sequencing, we recommended using DNBSEQ-G400RS Genetic Sequencer with DNBSEQ-G400RS High-throughput Sequencing Set (FCL PE150).

C. QTL-seq, genetic map construction and gene mapping

During QTL-seq analysis, Burrows-Wheeler Aligner (BWA) was used to map clean reads to reference genome. SAMtools were applied to convert aligned reads into SAM/BAM files. the GATK software was used to identify homologous SNPs in the two parents. The genome resequencing data of the 200 individuals in the F_2 population were filtered and mapped to the reference genome using the same method as in QTL-seq. SNPs were identified using an in-house pipeline in 'Sentieon Genomics' tools. False SNP loci were filtered using the GATK

software. The construction of the high-density genetic map and QTL analysis were conducted using the software HighMap and R/qtl.

D. RNA-seq Analysis

The RNA-seq reads were filtered and aligned to the reference genome using the software SOAPnuke and HISAT. The alignment of clean reads to the reference genome and the calculation of gene alignment rate were obtained using Bowtie2. The expression levels of genes and transcripts were calculated using the RSEM software. Differentially expressed genes (DEGs) were identified using the DEseq2 software. The Kyoto Encyclopedia of Genes and Genomes (KEGG) enrichment analysis was carried out using the software ClusterProfiler and ggplot.

Sample collection	Library preparation and sequencing	Bioinformatics analysis	Result analysis
a tall mutant ZS11-HP and a dwarf mutant sdw-e were conducted parental cross to construct F_2 segregating population. Two hundred individuals in the F_2 population were planted as genetic analysis strains. Two extreme bulks were constructed.	 MGIEasy Universal DNA Library Prep Set MGIEasy RNA Library Prep Set  DNBSEQ-G400 Genetic Sequencer	SAMtools Sentieon Genomics GATK HighMap R/qtl SOAPnuke HISAT Bowtie2 RSEM DEseq2 ClusterProfiler ggplot	Analysis of QTL-seq and development of a high-density genetic map, Genetic mapping, Analysis of gene expression levels

Results

Analysis of QTL-seq and development of a high-density genetic map

After the sequencing results of the two parents and two extreme bulks were aligned to reference genome, 936,889 polymorphic markers were obtained in the whole genome. These markers were utilized to calculate the allelic variation frequency of the two extreme bulks (SNP-index).

The 200 individuals in the F_2 population were subjected to WGS to develop the high-density genetic map and a total of 1,841.29 Gb clean data were gained. The average quality score of Q20 exceed 96.89%, the GC content was 37.20%, the average sequencing depth was 7.67 x, and finally a total of 677,649 SNPs were detected. Among the SNPs, 393,214 were employed to develop the genetic map (Table 1). A total of 4,323 bin markers were utilized for the genotyping of the 200 individuals. Eventually, a genetic map containing 19 linkage groups was successfully constructed with a total distance of 2,026.52 cM. The average genetic distance between neighboring markers was 0.47 cM. The longest linkage group was C02, reaching 175.64 cM, while the shortest was A01, only 62.59 cM (Figure 2).

Sample Name	Clean Reads	Clean Base	Read Length	Q20(%)	GC(%)
18ZT332-45A	68689280	10303392000	150	97.25	36.83
18ZT332-47A	65634172	9845125800	150	97.08	37.36
18ZT332-48A	68913330	10336999500	150	97.45	37.23
18ZT333-01A	65750152	9862522800	150	97.54	36.47
18ZT333-02A	71433436	10715015400	150	97.53	35.78
18ZT333-03A	63931548	9589732200	150	97.34	37.42
18ZT333-04A	56973952	8546092800	150	96.82	36.73
18ZT333-05A	62668678	9400301700	150	97.08	37.28
18ZT333-06A	61168774	9175316100	150	97.13	36.63
18ZT333-08A	54397162	8159574300	150	96.99	36.17
18ZT333-09A	54159854	8123978100	150	96.7	36.5
18ZT333-10A	63361192	9504178800	150	97	36.72
18ZT333-11A	61498332	9224749800	150	96.84	36.94
18ZT333-12A	60636214	9095432100	150	96.71	37.4
18ZT333-13A	65356366	9803454900	150	97.33	36.57
18ZT333-14A	74944962	11241744300	150	97.42	36.3
18ZT333-15A	65111186	9766677900	150	97.21	36.43
18ZT333-16A	67963740	10194561000	150	97.03	36.35
18ZT333-18A	62715008	9407251200	150	97.01	36.68
18ZT333-19A	60755716	9113357400	150	97.15	35.73

Table 1. Showing the resequencing quality of a subset of the 200 individuals from the F_2 segregating population.

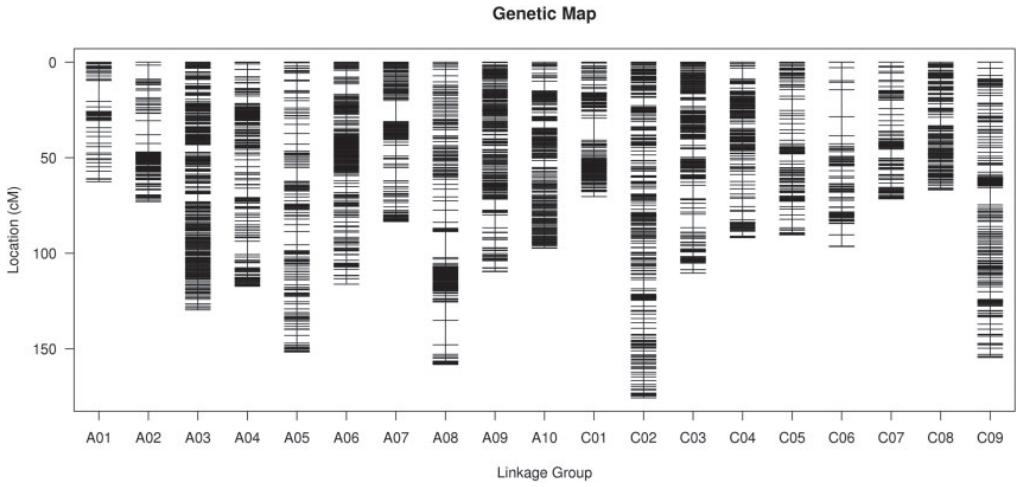

Figure 2. High-density map of *Brassica napus*.

QTL genetic mapping

The QTL interval of *Brassica napus* height was determined using interval mapping (IM) and composite interval mapping (CIM) of the R/qtl software. Combining the analysis results of QTL-seq, a major QTL regulating *Brassica napus* height was eventually identified on chromosome 10 (Figure 3).

Analysis of gene expression levels

A total of 47.72 Gb clean data were obtained after sequencing the transcriptome of the two parents (ZS11-HP and sdw-e). The average quality score of Q30 was above 95.48%. The transcriptome data can be considered reliable based on the analysis results of the R^2 value. A total of 11 upregulated and 14 downregulated candidate genes were identified within the determined QTL intervals. The analysis of WGS and QTL-seq data revealed the presence of 1,226 homozygous SNPs and InDels within the QTL intervals. Through the integration of the genetic variation and expression data of the two parents, researchers identified three candidate genes *BnaA10g08290D*, *BnaA10g09290D*, and *BnaA10g08230D*, which are believed to regulate the height of *Brassica napus*. These genes carried missense mutations between the two parents, and their expression levels were significantly different from each other (Figure 4).

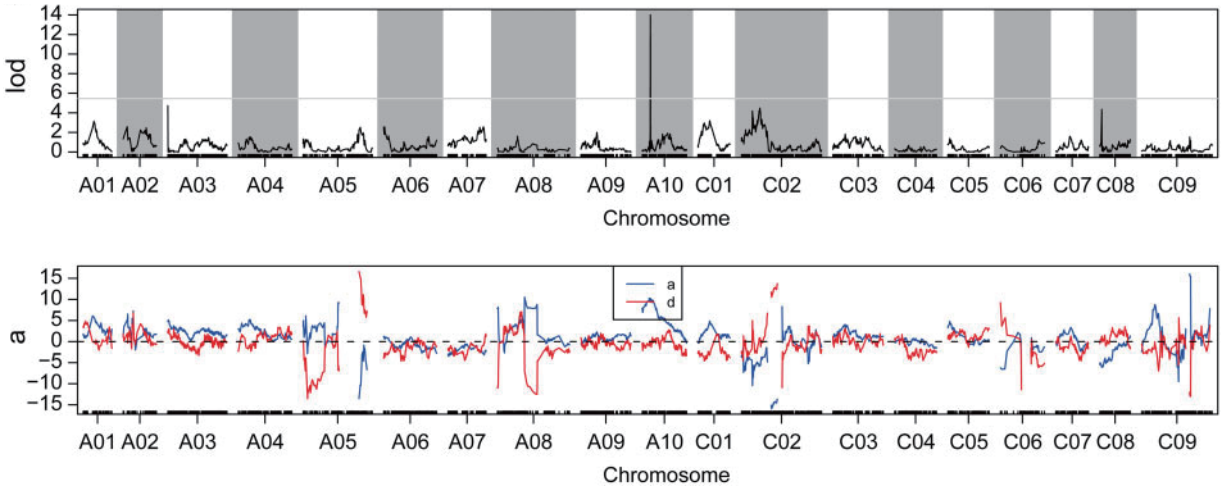

Figure 3. Results of QTL mapping.

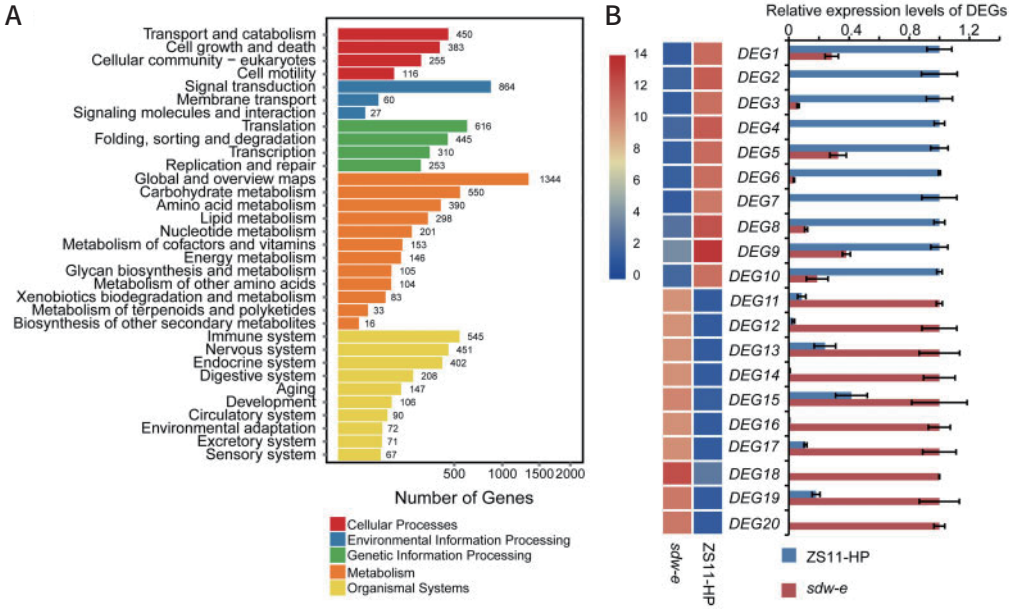

Figure 4. KEGG enrichment analysis and RT-qPCR verification of DEGs.

Summary

In this study, researchers conducted WGS for the 200 individuals in the *Brassica napus* F_2 segregating population and RNA-seq for specific parts of the two parents using the MGI sequencing platform. As a result, high-quality, highly-accurate sequencing data were obtained. After analyzing the data, researchers identified the major QTL responsible for controlling the height of *Brassica napus* on chromosome 10. By analyzing transcriptome data and genetic variations, the candidate genes were identified. This provides new molecular markers for marker-assisted breeding of *Brassica napus*.

The MGI DNBSEQ sequencing platform, which relies on its self-developed DNBSEQ™ technology, offers several advantages including high accuracy, low redundancy, low index hopping rate. The DNBSEQ genetic sequencer can be used to identify candidate genes related to dwarfing breeding, yield increasement, plant architecture enhancement in *Brassica napus*. This will facilitate molecular marker-assisted breeding of *Brassica napus*.

DNBSEQ-G400 Genetic Sequencer

References

1. Dong Z, Alam M K, Xie M, et al. Mapping of a major QTL controlling plant height using a high-density genetic map and QTL-seq methods based on whole-genome resequencing in Brassica napus. *G3 (Bethesda)*, 2021, 11(7):jkab118.
2. Khush G S. Green revolution: the way forward. *Nat Rev Genet*, 2001, 2(10): 815-22.
3. Bin Y I, Wei C, Chao-Zhi M A, et al. Mapping of quantitative trait loci for yield and yield components in Brassica napus L. *Acta Agronomica Sinica*, 2006.
4. Huang C, Yang M, Shao D, et al. Fine mapping of the BnUC2 locus related to leaf up-curling and plant semi-dwarfing in Brassica napus. *BMC Genomics*, 2020, 21(1): 530.
5. Salvi S, Tuberosa R. To clone or not to clone plant QTLs: present and future challenges. *Trends Plant Sci*, 2005, 10(6): 297-304.
6. Chalhoub B, Denoeud F, Liu S, et al. Plant genetics. Early allopolyploid evolution in the post-Neolithic Brassica napus oilseed genome. *Science*, 2014, 345(6199): 950-3.
7. Hua Y, Zhang D, Zhou T, et al. Transcriptomics-assisted quantitative trait locus fine mapping for the rapid identification of a nodulin 26-like intrinsic protein gene regulating boron efficiency in allotetraploid rapeseed. *Plant Cell Environ* , 2016, 39(7): 1601-18.

Recommended Ordering Information

Category	Product	Cat. NO.
Instruments	Genetic Sequencer DNBSEQ-G400RS	900-000170-000
	MGISP-100RS Automated Sample Preparation System	900-000206-00
	MGISP-960RS Automated Sample Preparation System	900-000146-00
Software	MegaBOLT Bioinformatics analysis accelerator	900-000555-00
Library Prep	MGIEasy Universal DNA Library Prep Set (16 RXN)	1000006985
	MGIEasy RNA Library Prep Set (16 RXN)	1000006383
Sequencing Reagents	DNBSEQ-G400RS High-throughput Sequencing Set (FCL PE150)	1000016952

MGI Tech Co.,Ltd

Building 11, Beishan Industrial Zone, Yantian District, Shenzhen, CHINA, 518083

Version: August 2023

+86-4000-688-114

en.mgi-tech.com

MGI-service@mgi-tech.com

Author: Jie Chen, Hao Zhang

Editor-in-Charge: Qiwei Wang

Reviewer: Yao Jiang

1. For StandardMPS and CoolMPS: Unless otherwise informed, StandardMPS and CoolMPS sequencing reagents, and sequencers for use with such reagents are not available in Germany, Spain, UK, Sweden, Belgium, Italy, Finland, Czech Republic, Switzerland, Portugal, Austria and Romania. Unless otherwise informed, StandardMPS sequencing reagents, and sequencers for use with such reagents are not available in Hong Kong. No purchase orders for StandardMPS products will be accepted in the USA until after January 1, 2023.
2. For HotMPS sequencers: This sequencer is only available in selected countries, and its software has been specially configured to be used in conjunction with MGI's HotMPS sequencing reagents exclusively.
3. For HotMPS reagents: This sequencing reagent is only available in selected countries.