Plant Science xxx (2015) xxx–xxx

Contents lists available at ScienceDirect

Plant Science

journal homepage: www.elsevier.com/locate/plantsci

A genome-wide association study of plant height and primary branch number in rapeseed (*Brassica napus*)

Feng Li[a,b], Biyun Chen[a], Kun Xu[a], Guizhen Gao[a], Guixin Yan[a], Jiangwei Qiao[a], Jun Li[a], Hao Li[a], Lixia Li[a], Xin Xiao[a], Tianyao Zhang[a], Takeshi Nishio[b], Xiaoming Wu[a,*]

[a] *Key Laboratory of Biology and Genetic Improvement of Oil Crops, Ministry of Agriculture, Oil Crop Research Institute, Chinese Academy of Agricultural Sciences, Wuhan 430062, China*
[b] *Graduate School of Agricultural Science, Tohoku University, 1-1 Tsutsumidori Amamiyamachi, Aoba-ku, Sendai, Miyagi 981-8555, Japan*

ARTICLE INFO

Article history:
Available online xxx

Keywords:
Brassica napus
Association mapping
SNP
Plant height
Primary branch number
Plant architecture

ABSTRACT

Crop plant architecture plays a highly important role in its agronomic performance. Plant height (PH) and primary branch number (PB) are two major factors that affect the plant architecture of rapeseed (*Brassica napus*). Previous studies have shown that these two traits are controlled by multiple quantitative trait loci (QTL); however, QTLs have not been delimited to regions less than 10 cM. Genome-wide association study (GWAS) is a highly efficient approach for identifying genetic loci controlling traits at relatively high resolution. In this study, variations in PH and PB of a panel of 472 rapeseed accessions that had previously been analyzed by a 60k SNP array were investigated for three consecutive years and studied by GWAS. Eight QTLs on chromosome A03, A05, A07 and C07 were identified for PH, and five QTLs on A01, A03, A07 and C07 were identified for PB. Although most QTLs have been detected in previous studies based on linkage analyses, the two QTLs of PH on A05 and the QTL of PB on C07 were novel. In the genomic regions close to the GWAS peaks, orthologs of the genes involved in flower development, phytohormone biosynthesis, metabolism and signaling in *Arabidopsis* were identified.

1. Introduction

Crop plant architecture is defined as the three-dimensional organization of the vegetative and reproductive organs. Plant architecture strongly affects light interception and photosynthesis in the canopy, playing an extremely important role for overall yield and harvest index [1,2]. In domesticated crops, breeding efforts have focused on plant architecture modification, i.e., "ideal-type breeding," which has been adopted to improve crop adaptability to different environments and to increase seed/fruit yield [3–8]. Notably, one of the great successes of the Green Revolution was development of the varieties with reduced height and more robust stems [4]. During modern breeding, plant architecture modifications aim to increase management and harvest convenience, to make favorable partition of carbon and nutrients between grains and the other parts and to enhance efficiency of fertilizer and water use [9]. Understanding the genetic basis of a crop's plant architecture will facilitate integration of plant architectural traits into the ideal plant architecture by molecular design breeding to improve agronomic performance [9,10].

Plant height and shoot branching are the major factors in plant architecture construction. Gibberellin (GA) is a phytohormone that is necessary for internode elongation in vegetative plants. Several GA biosynthesis- and metabolism-related genes have been identified affecting plant height in *Arabidopsis thaliana* [11], rice [12], and maize [13]. However, shoot branching is controlled by a more complicated network. In *A. thaliana*, genetic basis of shoot branching has been well studied via mutagenesis [14,15]. The integrated regulatory network that controls shoot branching includes flowering time genes [16], floral meristem identity genes [17,18], node patterning genes [15,19], and phytohormone (e.g., auxin, cytokinin, and strigolactones) biosynthesis, transport, and signaling genes [11,14,19]. The genetic pathways that determine *A. thaliana* inflorescence architecture are highly conserved in flowering plants [15,18,20]. Information from these genes may help identify candidate genes underlying phenotypic variation of plant architecture in other species.

Rapeseed (*Brassica napus* L., AACC, $2n = 38$) is the second most important oilseed crop in the world after soybean for the production of oil seeds and oil meals (protein source). After flowering induction, rapeseed generates racemose inflorescence bearing

* Corresponding author. Tel.: +86 27 8681 2906; fax: +86 27 8681 2906.
E-mail address: wuxm@oilcrops.cn (X. Wu).

http://dx.doi.org/10.1016/j.plantsci.2015.05.012
0168-9452/

flowers and secondary inflorescences similar to the closely related *A. thaliana*. Rapeseed plant architecture is characterized by plant height, branch number and distribution, and length of main inflorescence. These traits indirectly influence rapeseed cultivar yield by affecting the major yield-component trait, e.g., number of siliques per plant (PS) [21,22]. Plant height (PH) negatively correlates with PS, and taller plants leads to higher risk of lodging at late stages of rapeseed development [21]. The number of primary branches (PB) significantly positively correlates with PS [21]. Although plant architecture-associated traits respond to environmental conditions [23], they have high heritability [21,24]. In the past two decades, the genetic bases of these traits have been studied by quantitative trait loci (QTL) mapping using biparental cross-derived segregating populations [25–30]. However, these QTL studies typically localize QTLs to 10–20 cM intervals due to the limited number of recombination events that occurred during the construction of mapping populations.

Genome-wide association study (GWAS) takes full advantage of historic recombination events and is performed by scanning a genome using high-density DNA markers to identify genetic loci underlying traits of interest at a relatively high resolution [31,32]. Recent studies have shown that GWAS is a powerful tool to detect the genetic architecture of traits in many crops, and it can identify multiple related candidate genes [33–35]. In our previous study, we performed genotype analysis of an association panel with 472 accessions using a 60k *Brassica* Infinium® SNP array, the SNPs of which, were in silico mapped in the *B. napus* pseudomolecules created by Harper et al. [36]. We then performed a GWAS of *B. napus* seed weight and seed quality [35]. We identified several loci or candidate genes consistent with previous studies, indicating that our rapeseed panel is suitable for fine mapping complex traits. The draft *B. napus* genome sequence was recently published [37], which may provide more accurate genomic information for GWAS than the *B. napus* pseudomolecules. In the present study, the SNPs were in silico mapped in the *B. napus* genome sequence, and a GWAS of plant architecture-related traits, including PH and PB, was performed to uncover the genetic bases of these traits.

2. Materials and methods

2.1. Plant materials

An association population of rapeseed previously reported, including 472 accessions [35], was used as plant materials in this study. One representative plant of each accession was self-pollinated in spring of 2011, and the leaves of the representative plants were sampled to extract genomic DNA for SNP chip analysis. The self-pollinated seeds of each accession were used for trait investigation.

2.2. Experimental design and trait measurements

Trials were conducted at the experimental farm of the Oil Crops Research Institute of the Chinese Academy of Agricultural Sciences in Yangluo, Wuhan, China (114.51°E, 30.71°N). The rapeseed accessions were grown following a randomized complete block design with three replications in three growing periods from September in 2011 to May in 2014 (referred to as 2012, 2013, and 2014). The cultivation and management methods were described previously [21]. Briefly, the plants were grown in plots 1.67 m long and 2.40 m wide, with a 0.33 m spacing between rows and 0.15 m between plants within each row.

At maturity (in May), five plants from the center row of each plot were used for investigating their PH and PB. PH was measured as the height from the base of the stem to the tip of the main inflorescence. PB was measured as the number of primary branches arising from the main shoot. The mean value of the five plants in a plot represents the trait value of an accession grown in that plot.

2.3. Statistical analysis

For each trait, the variance components were estimated using the linear mixed effects "lmer" command of the lme4 package [38] in R software Version 3.11 for Windows, and the following linear model was used:

$$Y_{ij} = \mu + G_i + Y_j + GY_{ij} + \varepsilon_{ij}$$

where Y_{ij} is the trait measured, μ is the overall mean, G_i is the effect resulting from the ith genotype, Y_j is the effect resulting from the jth year, GY_{ij} is the effect resulting from genotype × year (environment) interaction, and ε_{ij} is the residual error (effect resulting from the experimental error). All effects were treated as random. The estimated genetic variance ($\delta_g{}^2$), the variance due to the genotype × year interaction ($\delta_{ge}{}^2$), and the residual error ($\delta_e{}^2$) were used for calculating broad-sense heritability as $H^2 = \delta_g{}^2/(\delta_g{}^2 + \delta_{ge}{}^2/n + \delta_e{}^2/nr)$, where n is the number of years and r is the number of replicates within a year.

Best linear unbiased predictors (BLUP) were estimated for each line for each trait based on the above-mentioned linear model using the lme4 package [38]. The BLUP values and single-year values for each genotype were used for the association analysis.

2.4. In silico mapping of SNPs

The probe sequences of the 26,841 high-quality SNPs previously selected [35] were used to perform a BlastN [39] search against *B. napus* genome sequences [37]. Only the top blast-hits with an *E*-value cut-off of 1E-15 against the *B. napus* genome sequences were considered. Furthermore, blast matches to multiple loci with the same *E*-value were excluded.

2.5. Population structure and Linkage disequilibrium (LD) analysis

We used PLINK [40] –indep-pairwise 50 5 0.8 to exclude the SNPs in high LD ($r^2 > 0.8$), leaving 7410 SNPs. These SNPs were used for principal component analysis (PCA) and relative kinship analysis. PCA was performed using the GCTA tool [41], and the top five principal components were used for constructing a P matrix. The relative kinship K matrix was constructed by the software package SPAGeDi [42]. LD was calculated using the software TASSEL 3.0 [43]. The overall LD decay in relation to physical distance was evaluated with the R package ggplot2 [44] using the modified nonlinear regression of r^2 with the equation $E[r^2] = 1/(1 + 4N_e c)$, where Ne is the effective population size and c is the physical position [45].

2.6. Genome-wide association analysis

TASSEL 5.0 software [43] was used for GWAS using the PCA model and PCA + K model. The PCA model was performed using a general linear model (GLM) with the following equation:

$$y = X\alpha + e.$$

The PCA + K model was performed using a mixed linear model (MLM) with the equation $y = X\alpha + K\mu + e$.

In these equations, y represents phenotype, X represents genotype, α is a vector containing fixed effects, including genetic marker and population structure (PCA), K is the relative kinship matrix, μ is a vector of random additive genetic effects, and e is the unobserved vector of random residual.

Please cite this article in press as: F. Li, et al., A genome-wide association study of plant height and primary branch number in rapeseed (*Brassica napus*), Plant Sci. (2015), http://dx.doi.org/10.1016/j.plantsci.2015.05.012