Plant Science 252 (2016) 388–399

Contents lists available at ScienceDirect

Plant Science

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

Identification of stable QTLs for seed oil content by combined linkage and association mapping in *Brassica napus*

Fengming Sun [1], Jing Liu [1], Wei Hua, Xingchao Sun, Xinfa Wang, Hanzhong Wang*

Oil Crops Research Institute of the Chinese Academy of Agricultural Sciences, Key Laboratory of Biology and Genetic Improvement of Oil Crops, Ministry of Agriculture, Wuhan 430062, PR China

ARTICLE INFO

Article history:
Received 27 May 2016
Received in revised form 1 September 2016
Accepted 6 September 2016
Available online 8 September 2016

Keywords:
Linkage mapping
Brassica napus
Quantitative trait locus
GWAS

ABSTRACT

Seed oil content is an important agricultural trait in rapeseed breeding. Although numerous quantitative trait locus (QTL) have been identified, most of them cannot be applied in practical breeding mainly due to environmental instability or large confidence intervals. The purpose of this study was to identify and validate high quality and more stable QTLs by combining linkage mapping and genome-wide association study (GWAS). For linkage mapping, we constructed two F_2 populations from crosses of high-oil content (~50%) lines 6F313 and 61616 with a low-oil content (~40%) line 51070. Two high density linkage maps spanned 1987 cM (1659 bins) and 1856 cM (1746 bins), respectively. For GWAS, we developed more than 34,000 high-quality SNP markers based on 227 accessions. Finally, 40 QTLs and 29 associations were established by linkage and association mapping in different environments. After merging the results, 32 consensus QTLs were obtained and 7 of them were identified by both mapping methods. Seven overlapping QTLs covered an average confidence interval of 183 kb and explained the phenotypic variation of 10.23 to 24.45%. We further developed allele-specific PCR primers to identify each of the seven QTLs. These stable QTLs should be useful in gene cloning and practical breeding application.

1. Introduction

Vegetable oils represent a major feedstock for food, animal feed, and chemical industries, as well as being a renewable source of energy [1,2]. With recent population growth and economic development, global consumption of vegetable oils has increased by more than 50% over the past decade [1]. Rapeseed (*Brassica napus* L.) is the third largest source of vegetable oil, and oil content continues to be an acknowledged major trading commodity and an important goal of oil crop breeding. As one of determinants of oil production, seed oil content is the subject of many current studies [3].

Oil content is typically a quantitative trait regulated by multiple genes. In the past decade, numerous QTLs for seed oil content have been identified [4–12]. Most studies used traditional methods to develop DNA markers, Ecka et al. [7] developed restriction fragment length polymorphism (RFLP) markers to construct genetic maps and mapped QTLs for erucic acid and seed oil content, Delourme et al. [5] used simple sequence repeat (SSR) markers to study the genetic control of oil content in *B. napus*, and Yan et al. [4] combined sequence-related amplified polymorphism (SRAP), SSR, amplified fragment length polymorphism (AFLP), and target region amplified polymorphism (TRAF) markers to perform a linkage analysis for seed oil content, seed hull content and seed coat color. The number of polymorphic DNA markers varied from 125 to 527, and the interval distance between adjacent markers ranged from 3.5 to 8.8 cM. Recent advances in genotyping technology have made possible the construction of high-density genetic maps. For example, using single nucleotide polymorphisms (SNPs), SSR, AFLP arrays, and kompetitive allele specific PCR (KASP) markers, Teh et al. (2016) constructed a high-density genetic map containing 1638 markers with average intervals of 2.0 cM, and anchored three oil

Abbreviations: QTL, quantitative trait locus; SNP, single nucleotide polymorphism; GWAS, genome-wide association study; RFLP, restriction fragment length polymorphism; RAPD, random amplified polymorphic DNA; AFLP, amplified fragment length polymorphism; CAPS, cleaved amplified polymorphism sequences; IFLP, intron fragment length polymorphism; SSR, sample sequence repeat; G1, 6F313 × 51070 $F_{2(2:3)}$ population; G2, 61616 × 1070 $F_{2(2:3)}$ population; PIC, polymorphism information contents; LD, linkage disequilibrium; K matrix, kinship matrix; CIM, composite interval mapping; MCMC, Markov Chain Monte Carlo; MAF, Minimum allele frequency.

* Corresponding author.
E-mail addresses: sunfengming@caas.cn (F. Sun), liujing@oilcrops.cn (J. Liu), huawei@oilcrops.cn (W. Hua), sunxingchao_lover@126.com (X. Sun), wangxf12@sina.com (X. Wang), wanghz@oilcrops.cn (H. Wang).

[1] These authors contributed equally to this work.

http://dx.doi.org/10.1016/j.plantsci.2016.09.001
0168-9452/

content related QTLs [13]. Generally, marker-assisted selection (MAS) should provide an important foundation for crop trait improvement. However, environmental variation can substantially reduce the applicability of QTLs, especially the seed oil content related QTLs in breeding programs. The impact of different environmental factors (e.g., temperature, light) made many QTLs invalid in different environments. For example, Chen et al. identified 27 QTLs for oil content across five different environments, but 23 were associated with only a single environment and 4 were identified in at least two [11]. In the Rapid × NSL96/25 DH population, 3 of 18 QTLs associated with oil content were reproducible across three trials [5]. Therefore, stable QTLs with small confidence intervals will be more valuable for molecular breeding.

Another alternative method for mapping QTL is genome-wide association study, originally designed to identify genes related to complex human diseases but now increasingly used in plant genetics. In recent years, GWAS has been successfully applied to crop plants such as rice and maize [14–16]. For rapeseed, several previous studies were conducted to dissect linkage disequilibrium [17–22]; however, highly homologous subgenomes [23] and earlier technology limitations hindered the development of molecular markers at a sufficient density. SNP array technology can result in higher-density markers at a genome-wide level. Delourme et al. [24] developed more than 4300 SNP markers from 313 inbred lines and Li et al. [25] obtained a high density 26,841 SNP marker set from 472 accessions. Because the two subgenomes of *B. napus* (AACC, 2n = 38; formed ~7500 years ago) are highly conserved with the diploid parental genomes of *B. rapa* (AA, 2n = 20) and *B. oleracea* (CC, 2n = 18) [26], the parental species were used to supersede *B. napus* as the reference genomes in previous association studies, such that of Li et al. who successfully identified a QTL related to seed oil content on chromosome A08 [25]. Now, the assembled entire genome of *B. napus* (cv. Darmor-*bzh*) has been completed and published [26], providing an opportunity to conduct GWAS using *B. napus* itself. Although association studies have advantages in QTL mapping, false positive results can be derived from population structure and individual relationships [27]. Therefore, it is necessary to verify the results of GWAS with other methods such as linkage mapping. Liu et al. [28] applied GWAS to *B. napus* to study seed oil content and identified a novel locus on the chromosome A05 that was successfully validated by linkage mapping.

In this study, we aim to identify and establish more reliable QTLs by combining linkage mapping and GWAS. In order to include more beneficial alleles in linkage mapping, we selected two parental lines with special genetic profiles for oil accumulation to constructed F_2 segregation populations [29]. Meanwhile, a germplasm set of 227 accessions were used to perform the association study. To develop more SNP markers and facilitate comparison, a 60K *Brassica* Infinium SNP array was used to genotype all individuals in this study. Finally, we identified seven overlapping QTLs which could be useful for further gene cloning and breeding application.

2. Materials and methods

2.1. Field trials, measurement of seed oil content, and genotyping for individuals

Two high-oil content (HO) lines, 6F313 and 61616 (seed oil contents exceeding 50%) with different genetic backgrounds [29–31], were respectively crossed with a low oil content (LO) line, 51070. Two F_2 populations of 105 individuals were produced. Field experiments including the parents and segregating populations (2010: F_2, 2011 $F_{2:3}$ and 2012 $F_{2:3}$) were conducted in a thrice replicated, randomized complete block design, in Yangluo, Hubei. The $F_{2:3}$ population seeds were sown by hand in double row plots, and field management was followed according to standard agricultural practices. Each row was 2.5 m long, with 40 cm between rows and 20 cm between plants. At maturity, 6–10 individual rapeseed plants were harvested from each plot. The seed oil content was measured using nuclear magnetic resonance (NMR PQ001; Niumag, Wuhan, Hubei, China). The detection range of seed weight with NMR PQ001 was about 1.0–1.2 g. Seed oil content was also evaluated for parents (6F313, 61616 and 51070) in years 2010, 2011 and 2012. Total of 227 germplasm accessions were collected from different origins, including 122 accessions from middle of the Yangtze River of China, 46 from upper reaches of the Yangtze River of China, 36 from lower reaches of Yangtze River of China, 10 from Europe, 8 from Australia and 5 from other places or unknown origins, were planted in 2013 and 2014 (Wuhan, Hubei) and their oil content was measured by NMR PQ001 (Table S1). All individuals from the segregating populations and germplasm lines were genotyped by using the same 60K *Brassica* Infinium SNP array developed by the international *Brassica* Illumina SNP consortium.

2.2. Statistical analysis for broad-sense heritability

The Shapiro–Wilk normality test was used to evaluate the frequency distributions of trait data, with the null hypothesis being that data were taken from a normal distribution. The correlation coefficient of oil content between each two trials was calculated with a Pearson's correlation test. The broad-sense heritability was calculated as $h^2 = {\sigma_g}^2/({\sigma_g}^2 + {\sigma_{ge}}^2/n + {\sigma_e}^2/nr)$, where h^2 represents broad-sense heritability, ${\sigma_g}^2$ represents the genetic variance, ${\sigma_{ge}}^2$ represents the interaction variance of genotypes with environment, ${\sigma_e}^2$ is the error variance, n is the number of environments and r is the number of replications. The elements in the formula were calculated by SAS ANOVA procedure.

2.3. Construction of genetic linkage maps

Genetic linkage maps for the 6F313 × 51070 (G1) and 61616 × 51070 (G2) F_2 populations were constructed using JoinMap version 4.1 [32]. First, we converted the genotype file to a format that could be processed by JoinMap using a custom PERL script, and then merged successive markers with the same genotype into a single bin. Three steps were then used to construct the linkage maps: (1) filtering, where the segregation rate of each marker in the F_2 population was evaluated by a chi-square goodness-of-fit test for a ratio of 1:2:1 or 1:3, and where the missing rate was less than 20%; (2) grouping, where the initial LOD (likelihood of odds) score for grouping was set to 20 and 19 for G1 and G2, respectively; (3) sorting, where the order and linkage distance of markers within each linkage group was calculated using maximum likelihood methods, and all genetic distances were expressed in centiMorgan (cM) units using the Kosambi function (Kosambi, 1944). Some JoinMap Monte Carlo ML parameters were altered to fit the data: chain length was 1000, initial acceptance probability was 0.15, cooling control parameter was 0.001, stopping after 10,000 chains without improvement, length of the burn-in was 20,000, number of Monte Carlo EM cycles was 12, chain length per Monte Carlo EM cycle was 1000, and sampling period for recombination frequency matrix samples was 5. For assigning the linkage groups to pseudo-chromosomes, we queried the SNP probes against the *Brassica napus* genome (Darmor-*bzh*) using BLAST [33], and ranked the results based on the total number of unique markers matched.

2.4. Detection of syntenic blocks between the physical genome and genetic linkage maps

To evaluate the quality of genetic maps, we selected uniquely mapped bins (randomly selecting a single marker to represent a