Oil content related candidate gene mapping in oilseed rape

ZHAO Jianyi[1], NI Xiyuan[1], HUANG Ruizhi[2], XU Fei[3], LEI Weixia[1]

[1] *Institute of Crop Science, Zhejiang Academy of Agricultural Sciences, 310021, Hangzhou, China*
[2] *Institute of Virology and Biotechnology, Zhejiang Academy of Agricultural Sciences, 310021, Hangzhou, China*
[3] *Institute of Bioinformatics, Zhejiang University, Hangzhou, China Email: jianyi32000@yahoo.com.cn*

Abstract

The SG-DH map was updated with combined 42 BBSRC microsatellites loci, 31 *Brassica* EST derived markers and 15 gene-specific marker loci from 12 candidate genes, which directly or indirectly related to the oil synthesis in oilseed rape. In total, the new map included 216 marker loci with total distance of 1979 cM. By using 112 universal SSR markers, the SG map was aligned with other published maps. Except N7 and N18, all the linkage groups are comparable with recently published three maps. Three candidate genes (LPAAT, 4CL and LIM) and two *Brassica* EST-STS markers were observed in flanking or closely to the regions of the five most significant oil QTL. LPAAT was mapped near the region of the strongest oil QTL. Two lignin synthesis related genes 4CL and LIM were linked to the QTL on N13 and N18. The sequence alignment of LPAAT showed that 98.7% homology was revealed in coding region between parents and published Bn-cDNA. Between parents, except a six continued nucleotides deletion, 8 and 6 SNPs were detected in intron and extron regions, respectively and caused two amino acid mutations. The interest was that all the mutations were occurred in beneficial QTL allele from 'Gaoyou'. No more information to show the relationship between LPAAT and oil content so far, however, further investigation on this gene might be continued and more oil related candidate genes in this QTL region will be surveyed in the coming days.

Key words: Oilseed rape, Oil content, Candidate gene, Map update, QTL mapping.

Introduction

The main value of oilseed rape is linked to the oil content of the harvested seeds, either for human nutrition or industrial supply. However, because of the complex gene control system, the oil content in rapeseeds was difficult to be effectively improved by conventional breeding procedures. In the past few years, several QTL for oil content have been identified in *Brassica napus* (Ecke et al. 1995; Burns et al. 2003; Zhao et al. 2005, 2006). More recently, Delourme et al.(2006) reported a mapping study with two genetic backgrounds of DH populations and compared the oil QTL with reported in other 4 publications. Qiu et al. (2006) identified 6 QTL located on separate genomic regions over environments and aligned them with the *A. thaliana* genome to look for candidate genes of controlling oil content. In spite of several mapping analysis for oil content in oilseed rape, little is still known on metabolism pathways that genes interact with total seed oil content.

In the present paper we report: i) the updated SG map (Zhao et al. 2005) with additional 88 marker loci including newly developed 31 locus-specific sequence-tagged site (STS) derived from *Brassica* EST sequences and 15 STS or CAPS (cleaved amplified polymorphic sequences) markers for 12 genes encoding enzymes involved in oil, protein and lignin content synthesize pathways, ii) their genetic mapping in SG population and their comparison with mapped oil QTL, iii) sequence alignment of LPAAT locus on N7 among parents and published Bn-cDNA. The aim of our study was to increase the marker density and start developing a functional gene map on the framework of published SG map (Zhao et al. 2005), in which genes are related directly or indirectly to oil content accumulation; to compare these genes with mapped oil QTL and to search for candidate genes, which putatively related to the oil content in rapeseed.

Material and Methods

Mapping population, phenotypic data and QTL mapping: The segregation DH population from the cross 'Sollux'× 'Gaoyou'was obtained as described in Zhao et al. (2005). It consisted of 282 DH lines and was named SG population. Phenotypic data was from six experiments, 4 of them were previously analyzed (Zhao et al. 2005) and another 2 trails were conducted in Hangzhou, Chian in the later two years of 2003-2005. QTL Mapper V1.0 (Wang et al. 1999) was used for mapping oil QTL choosing' Map Main-effect QTLs' model.

Genetic markers: Public SSR primer pair sequences were obtained from *Brassica* database, UK (http://www.brassica.bbsrc.ac.uk/BrassicaDB/). Markers prefixed ZAAS are STSs developed from *Brassica* EST sequence database (ftp://149.155.100.41/pub/brassica/BrassicaDB-DNA) in our lab. After removing the vectors and sequence recombination, consensus sequences were aligned to the *Arabidopsis thaliana* genome ATH1v5(http://www.tigr.org). Primer pairs were then designed using the software PRIMER 3 (Rozen and Skaletsky 1998) and tested by e-PCR verifying them both in *Brassica* and *Arabidopsis* genomes. In total around 300 pairs of public SSR and 150 EST derived primer pairs were used for amplifying PCR products with'Sollux','Gaoyou'and F_1 plants. In addition, 30 candidate genes were selected either from the *Arabidopsis* Lipid Gene Database (http://www.plantbiology.msu.edu/lipids/genesurvey) or kindly provided by Institute of Virology and Biotechnology, Zhejiang Academy of Agricultural Sciences, with a special focus on the genes involved in fatty acid, triacylglycerol, protein and lignin biosynthesis pathways. Sequences from *Brassica* database were aligned to these genes

then primer pairs were designed based on mRNA , full gene or ESTs when homologous sequences were available, otherwise, the primer pairs were produced from *Arabidopsis* gene or mRNA sequences. Primers were designed where possible from coding regions of genes flanking introns, avoiding UTRs. Four to seven primer pairs per gene (or mRNA) were designed, depending on the length of deduced sequences. Four restriction enzymes Hint III, EcoR I, BamH I and Ras I were tried to digest the PCR products around 1kb fragment, for which no polymorphisms could be detected by STS markers. For these gene-specific markers, the abbreviation of genes were directly marker names used on map.

PCR condition, product visualization: PCR reactions were performed following the protocol described in Zhao et al (2005). Amplification products were separated by electrophoresis in 6% non-denaturing polyacrylamide gels and visualized by a rapid silver-staining.

Cloning and sequencing: The locus specific DNA fragment of LPAAT on N7 in two parents were cloned into pMD18-T vector from TaKaRa Bio-Company. Sequencing of the positive clones was performed in the Shanghai Sangon Sequencing Center (Shanghai, China) with three clones for each parent. Further, the full-length DNA sequences were also amplified using the gene specific primers according to the published *Brassica* napus LPAAT mRNA (Z95637).

Results

Map update and functional mapping: 42 marker loci from 39 BBSRC SSR primer pairs, 31 STS loci from 28 EST derived primers were firstly added to the previously developed SG map (Zhao et al. 2005). They were evenly arranged in 17 and 14 linkage groups, respectively. Recently, we tried to develop the oil related candidate gene markers, starting with STS or CAPS. From selected 30 candidate genes, 15 loci of 12 genes had been assigned into 10 linkage groups. The information of mapped candidate genes was described in table 1. The total number of marker loci on the updated SG map (Fig 1) was increased to 216. Using 42 newly added SSR together with previously mapped 70 SSR markers designated CB, BRAS, MR and MD, the SG map was aligned with newly reported maps (Piguemal et al. 2005; Lowe et al. 2004 and Qiu et al. 2006). 22, 13 and 24 common markers and conserved marker order were identified between SG and F2 consensus map (Piguemal et al. 2005), between SG and N-0-61-9 (Lowe et al. 2004) and between SG and TN (Personal communication), matching 14, 13 and 13 pairs of linkage groups, respectively.

Table 1. The description of oil related candidate genes mapped on the SG_DH map

Biological function	Gene function	Primer based sequence	Primer derived from[a]	Name of gene loci	Marker assay	N
Fatty acid synthesis	Acetyl-CoA carboxylase	gi|32515088	Bn-mRNA	ACC(1-2)	CAPS, Rsa I	5
	Biotin carboxylase	gi|32497337	Bn-mRNA	CAC2(1)	Size	6
		gi|32497337	Bn-mRNA	CAC2(3-4)	CAPS, Rsa I	9
	Ketoacyl-ACP Synthase I	At5g46290	At- gene	KAS I(1)b	Size	17
	Ketoacyl-ACP Synthase III	gi|79320514	At-mRNA	KAS III(3)	Size	19
	Pyruvate Dehydrogenase	gi|42562292	At-mRNA	IAR4 (3)	Size	15
Synthesis of mb. lipids in endomembrane system	2-Lysophosphatidate Acyltransferase	gi|56835661	Bn-mRNA	LPAAT(1)a	Size	7
		gi|56835661	Bn-mRNA	LPAAT(1)b	Size	9
		gi|56835661	Bn-mRNA	LPAAT(2)	Size	9
Synthesis of membrane lipids	Digalactosyldiacylglycerol Synthase	Contig69075	Bn-EST	DGD2(1)	Size	19
Ratio of the protein/lipid content	Phosphoenolpyruvate carboxylase	gi|507807	Bn- gene	PEPCase (1-2)a	CAPS, Ras I	4
Lignin Synthesis	Cinnamoyl-CoA reductase	gi|32516766	Bn-mRNA	CAR(1-2)	CAPS, EcoR	9
	4-coumaroyl-CoA synthase	gi|32505783	Bn-mRNA	4CL(2)	Size	13
	Phenylalanine ammonia-lyase	gi|79507155	At-mRNA	PAL3 (2)	Size	4
Lignin transcript factor	Brassica napus LIM domain protein	gi|78927022	Bn-mRNA	LIM (1)	Size	18

[a] Bn-EST: EST sequence from B. *napus*; At-gene: A. *thaliana* gene; At-mRNA: A. *thaliana* mRNA , Bn-mRNA: B. *napus* mRNA

[b] Number in bracket indicate the primer order. Each candidate sequence was divided into more than 3 connected fragments for primer design (5′-3′) depending on the length of sequence. Letters indicate specific locus when polymorphic loci were more than 2 of one primer pair

Oil QTL and candidate genes: Table 2 shows the five consistent oil QTL identified over 6 environments and their positions on genomic regions. The QTL on N7 demonstrated the strongest effect and was detected over all

6 environments. This QTL was mapped between two functional markers, EST-STS marker ZAAS176 and a candidate gene loci LPAAT(1)a, with 26.2cM distance in between. Two QTL on N13 and N18 linked with genes related to or regulated Lignin biosynthesis. The former was identified particularly at the environment of Hangzhou, China while the later was detected at four of six locations. In the rest of two QTL on N1 and N12, which Sollux alleles contributed to oil content, no specific oil related candidate genes were found nearby, but EST derived STS markers were located in the QTL regions, which may also help to look for the candidate genes.

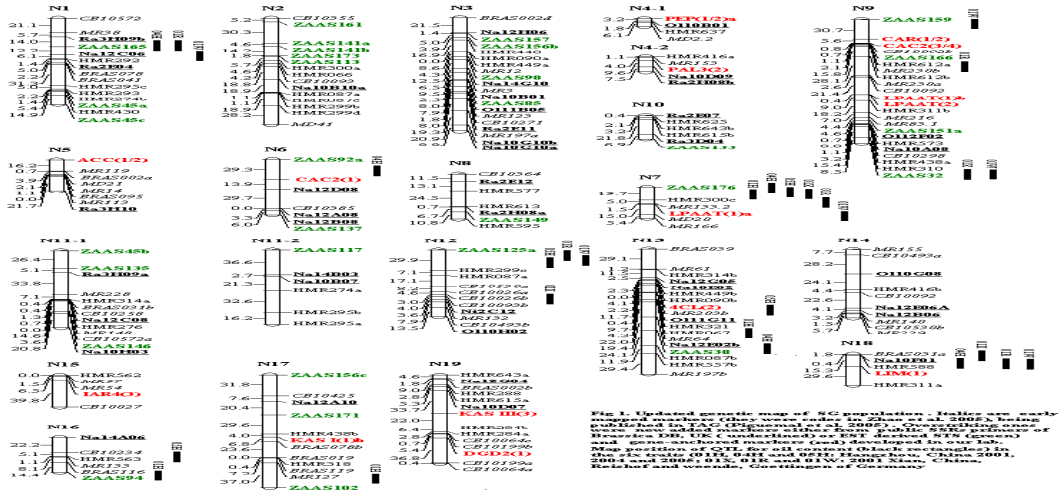

Table2 Five most significant QTL regions of oil content and the linked candidate genes

Flanking markers of QTL region (cM)	Location	Site[a]	LOD	Additivity[b]	R^{2c}
N1					
Ra3H09b-**ZAAS165**-HMR292	Hangzhou 2004	4.0	3.53**	0.41	3.1
32.3 cM	Reishof 2001	14.0	8.17***	0.66	6.5
	Weende 2001	32.3	6.30***	0.51	6.0
N7					
ZAAS176-LPAAT(1)a	Hangzhou 2001	4.0	9.7***	-0.81	15.8
26.2 cM	Hangzhou 2004	0.0	7.7***	-0.69	8.5
	Hangzhou 2005	4.0	15.6***	-1.08	21.5
	Xian 2001	8.0	5.78***	-0.50	9.1
	Reishof 2001	21.7	5.69***	-0.70	7.4
	Weende 2001	26.0	4.65***	-0.68	10.6
N12					
ZAAS125a-HMR299c	Hangzhou 2005	18.0	3.63**	0.66	8.0
29.9 cM	Reishof 2001	4.0	6.58**	0.71	7.5
	Weende 2001	10.0	2.34*	0.52	6.5
N13					
4CL(2)-ZAAS38	Hangzhou 2001	24.0	3.93***	-0.46	5.1
69.7 cM	Hangzhou 2004	64.0	4.51***	-0.60	6.5
	Hangzhou 2005	2.0	3.68***	-0.46	3.9
N18					
BRAS031a-**LIM(1)**	Hangzhou 2004	1.8	3.67***	-0.49	4.4
17.7 cM	Xian 2001	0.0	2.98***	-0.36	4.7
	Reishof 2001	1.8	3.17**	-0.57	4.9
	Weende 2001	1.8	4.70**	-0.49	5.5

[a] Distance of the QTL from the first marker of the indicated interval.
[b] Additive effect is the phenotypic effect due to the substitution of a Gaoyou allele by an allele of Sollux.
[c] R^2 is the percentage of variation explained by each QTL. $*P < 0.05$; $**P < 0.01$; $***P < 0.001$

Analysis of the cloned LPAAT gene: The N7-LPAAT locus polymorphic fragments of parents were cloned and sequenced (data not show). The results showed that thpolymorphism was caused by a (six continued nucleotides) 6bp deletion, located 12bp before the first intro/extron junction of "Gaoyou"allele. A locus specific primer pair flanking the deletion was further designed and tested in population. The result showed more clear, codominant and exactly the same fragment segregation pattern as mapped marker LPAAT(1)a (Fig 2). On the basis of the published LPAAT cDNA sequence, we designed a primer pair to amplify the full gene of two parents. High degree of homology had been found in corresponding fragments and full gene between two parents (>99%). Further, 98.7% identity was observed in coding region between parents and published Bn-cDNA. Alignment of full gene sequence between two parents demonstrated that, except one 6bp deletion, only 8 SNPs were detected in intron part while 6 SNPs were identified in extron regions, two of which could cause amino acid mutation. The point was that all the sequence mutations were occurred in the beneficial QTL allele from 'Gaoyou'.However, these results were very preliminary only from two replication sequence analysis and need to be further validated.

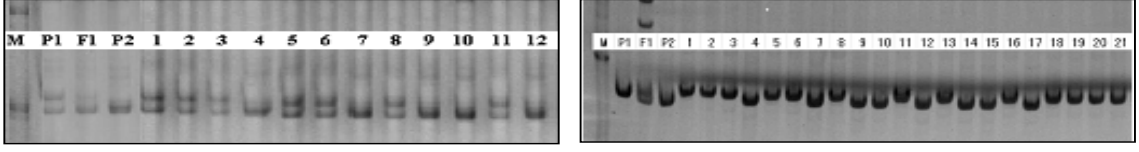

Fig 2. Segregation of STS locus detected by primer pair LPAAT(1) (A) and locus-specific primer pair directly flanking the 6bp deletion (B)

Discussion

The SG mapping population was advantaged with strong genetically and phenotypically diversity and larger segregation for oil content (40%-56%) (Zhao et al. 2005). However, small amount of marker loci and unpublished marker information made it difficult to compare the map and mapping results with other publications. The present SG map consisted of 216 marker loci in total, including newly developed 46 functional markers, which might be help to look for candidate genes when comparing them with mapped QTL. From the information of the Institute of Agronomy and Plant Breeding, University of Goettingen, Germany, previously mapped 70 SSR loci with code HMR, MR or MD were able to match the published markers (Piguemal et al. 2005). Together with 42 newly added BBSRC SSR loci, the new SG map is comparable with other maps (Piguemal et al. 2005; Lowe et al. 2004; Qiu et al. 2006; Delourme et al. 2006) in all linkage groups except N7 and N18, which no common markers shared in corresponding linkage groups due to limited universal markers. However, two significant oil QTL were located, respectively, on N7 and N18, therefore, to increase the marker density in these regions might be interesting using reference or gene-anchored markers.

The functional map allows to verify mapped QTL or mutations, if linked genes are good candidates. In this study, five most significant oil QTL regions were found either flanking or close to oil content related candidate genes or functional EST-STS markers. Although the gap between functional markers or the distance to the novel QTL is rather large, searching for more closed candidate genes by chromosome walking might be possible. The strongest oil QTL on N7 is located nearby an important candidate gene LPAAT, which catalyse the "2-Lysophosphatidate Acyltransferase", to form Diacylglycerol. In present study, no strong information supported the relationship between LPAAT and oil content in rapeseed. However, with flanking markers, the QTL-NILs are being developed and the cDNA of the gene will be sequenced in different organs of parent plants. On the other hand, inspection of the *Arabidopsis* lipid gene database in the region underlying QTL-N7 revealed several other genes that are involved in lipid metabolism as Diacylglycerol Acyltransferase, Ketoacyl-CoA Synthase, ER Phosphatidylglycerophosphate Synthase, Enoyl-CoA Reductase and so on, they are might be also candidate genes for working with in the coming days.

References

Burns MJ, Barnes SR, Bowman JG, Clarke MHE, Werner CP, Kearsey MJ (2003) QTL analysis of an intervarietal set of substitution lines in *Brassica napus*. I. Seed oil content and fatty acid composition. Heredity 90:39–48

Delourme R. , Falentin Huteau C, V. et al. (2006), Genetic control of oil content in oilseed rape (*Brassica napus* L.) Theor. Appl Genet. DOI 10.1007/s00122-006-0386-z

Ecke W, Uzunova M, Weißleder K (1995) Mapping the genome of rapeseed (*Brassica napus* L.). II. Localization of genes controlling erucic acid systhesis and seed oil content. Theor Appl Genet 91:972–977

Piquemal J, Cinquin E, Couton F, Rondeau C, Seignoret E, Doucet I, Perret D, Villeger MJ, Vincourt P, Blanchard P (2005)

Construction of an oilseed rape (*Brassica napus* L.) genetic map with SSR markers. Theor Appl Genet 111:1514–1523

Qiu D, Morgan C, Shi J, Long Y, Liu J, Li R, Zhuang X, Wang Y, Tan X, Dietrich E,.. J. Meng et al.(2006) A comparative linkage map of oilseed rape and its use for QTL analysis of seed oil and erucic acid content

Wang DL, Zhu J, Li ZK, Paterson AH (1999) Mapping of QTL with epistatic eVects and QTL £ environment interactions by mixed model approaches. Theor Appl Genet 99:1255 - 1264

Zhao J, Becker Heiko C, Zhang D, Zhang Y, Ecke Wolfgang (2005) Oil content in a European–Chinese rapeseed population: QTL with additive and epistatic effects and their genotype–environment interactions. Crop Sci 45:51–59

Zhao J , Becker HC, Zhang D, Zhang Y, Ecke W (2006) Conditional QTL mapping of oil content in rapeseed with respect to protein content and traits related to plant development and grain yield. Theor Appl Genet 113:33–38