Theor Appl Genet (2012) 125:715–729
DOI 10.1007/s00122-012-1863-1

ORIGINAL PAPER

Identification of *FAD2* and *FAD3* genes in *Brassica napus* genome and development of allele-specific markers for high oleic and low linolenic acid contents

Qingyong Yang · Chuchuan Fan · Zhenhua Guo · Jie Qin · Jianzhong Wu · Qingyuan Li · Tingdong Fu · Yongming Zhou

Received: 7 November 2011 / Accepted: 27 March 2012 / Published online: 26 April 2012
© Springer-Verlag 2012

Abstract Modification of oleic acid (C18:1) and linolenic acid (C18:3) contents in seeds is one of the major goals for quality breeding after removal of erucic acid in oilseed rape (*Brassica napus*). The fatty acid desaturase genes *FAD2* and *FAD3* have been shown as the major genes for the control of C18:1 and C18:3 contents. However, the genome structure and locus distributions of the two gene families in amphidiploid *B. napus* are still not completely understood to date. In the present study, all copies of *FAD2* and *FAD3* genes in the A- and C-genome of *B. napus* and its two diploid progenitor species, *Brassica rapa* and *Brassica oleracea*, were identified through bioinformatic analysis and extensive molecular cloning. Two *FAD2* genes exist in *B. rapa* and *B. oleracea*, and four copies of *FAD2* genes exist in *B. napus*. Three and six copies of *FAD3* genes were identified in diploid species and amphidiploid species, respectively. The genetic control of high C18:1 and low C18:3 contents in a double haploid population was investigated through mapping of the quantitative trait loci (QTL) for the traits and the molecular cloning of the underlying genes. One major QTL of *BnaA.FAD2.a* located on A5 chromosome was responsible for the high C18:1 content. A deleted mutation in the *BnaA.FAD2.a* locus was uncovered, which represented a previously unidentified allele for the high oleic variation in *B. napus* species. Two major QTLs on A4 and C4 chromosomes were found to be responsible for the low C18:3 content in the DH population as well as in SW Hickory. Furthermore, several single base pair changes in *BnaA.FAD3.b* and *BnaC.FAD3.b* were identified to cause the phenotype of low C18:3 content. Based on the results of genetic mapping and identified sequences, allele-specific markers were developed for *FAD2* and *FAD3* genes. Particularly, single-nucleotide amplified polymorphisms markers for *FAD*3 alleles were demonstrated to be a reliable type of SNP markers for unambiguous identification of genotypes with different content of C18:3 in amphidiploid *B. napus*.

Communicated by R. Visser.

Electronic supplementary material The online version of this article (doi:10.1007/s00122-012-1863-1) contains supplementary material, which is available to authorized users.

Q. Yang · C. Fan · Z. Guo · J. Qin · J. Wu · Q. Li · T. Fu · Y. Zhou (✉)
National Key Laboratory of Crop Genetic Improvement, Huazhong Agricultural University, Wuhan 430070, China
e-mail: ymzhou@mail.hzau.edu.cn

Introduction

Oilseed rape (*Brassica napus* L.) is one of the most important oil crops worldwide. The edible and processing quality of rapeseed oil is mainly determined by the fatty acid composition of the triglycerol lipid in its seeds. In particular, the proportions of the three major unsaturated fatty acids, oleic acid (C18:1), linoleic acid (C18:2) and linolenic acid (C18:3) in seeds are important for the nutritional and processing applications of oilseed rape. High C18:1 content in rapeseed oil is desirable for its long shelf life and stability in high temperature during deep frying (Warner and Knowlton 1997; Matthäus 2006). In addition, high oleic acid oil is also an ideal raw material for biodiesel production (Pinzi et al. 2009). On the other hand, linolenic acid is a highly unsaturated fatty acid; it is sensitive to oxidation in air and therefore unstable during frying. Therefore, a reduced C18:3 level in rapeseed oil is beneficial for storage and extending its shelf life (Scarth and Tang 2006). One of the major goals for quality breeding at present is to increase the C18:1 content (to >75 %) and reduce the C18:3 content

(to <3 %) in the seeds of oilseed rape cultivars (Scarth and Tang 2006; Wittkop et al. 2009).

Mutants with reduced C18:3 content of 3–5 % in *B. napus* seeds were first obtained by EMS mutation (Röbbelen and Nitsch 1975; Rakow 1973). The mutated materials with altered fatty acid profiles were widely used as a source of low C18:3 breeding and genetic studies. The first low C18:3 canola spring cultivar with 3 % C18:3 content, Stellar, was developed by crossing a low C18:3 germplasm M11 with a high C18:3 cultivar Regent (Scarth et al. 1988).

A molecular marker based on random amplified polymorphic DNA (RAPD) for C18:3 content was identified. The marker accounted for 26.5 % of the genetic variation of C18:3 content in an F_2 population derived from a cross between the cultivar Duplo and a low linolenic acid line (Hu et al. 1995). Two major QTLs, which explained 24 and 30.7 % of the total phenotypic variation of the C18:3 content, respectively, were determined with RAPD analysis in a double haploid (DH) population derived from a cross between Stellar and a high C18:3 cultivar Drakkar (Jourdren et al. 1996b). Similar results were obtained in other DH populations with the same low linolenic source (Rajcan et al. 1999; Thormann et al. 1996). The two loci were shown to correspond with the fatty acid desaturase 3 (*FAD3*) copies located in the A- and C-genome in *B. napus*, respectively (Barret et al. 1999; Jourdren et al. 1996a). The *FAD3* gene encodes an endoplasmic delta-15 linoleate desaturase that catalyzes the desaturation reaction of linoleic acid and the formation of linolenic acid (Arondel et al. 1992). The mutations in *FAD3* genes in Stellar could be detected by gene-specific markers, although the exact mutation sites were not known (Jourdren et al. 1996a). More recently, a single G-to-A mutation of the *FAD3* gene was identified to produce a reduced C18:3 content in DMS100, a high oleic and low linolenic canola line (Hu et al. 2006). Two major QTLs for C18:3 content were mapped on A4 and C4. Two point mutations in both *FAD3* loci located at A- and C-genome were reported in a low C18:3 winter rapeseed line DH219/05 (Mikolajczyk et al. 2010). The mutation in the allele of *FAD3* located at the C-genome of DH219/05 was exactly the same as the one in DMS100 (Hu et al. 2006; Mikolajczyk et al. 2010).

High oleic acid breeding came to the attention in oilseed rape breeding in the late 1980s. So far, studies on high oleic mutants in *B. napus* are limited. A QTL for C18:1 in spring turnip rape was identified in an F_2 population derived from the line Jo4002 (with 42 % of C18:1) and a high C18:1 individual Jo4072 (with 69 % of C18:1). A co-dominant RAPD marker was developed for the QTL (Tanhuanpää et al. 1996). The QTL was later found to correspond with the *FAD2* gene, in which there was a single base difference between the wild type and the high C18:1 allele. Such a difference allowed for the development of an allele-specific marker for high C18:1 for this material (Tanhuanpää et al. 1998). Through EMS mutation, two high C18:1winter mutant lines, 7488 and 19661, were developed from winter cultivar Wotan and the mutants had about 15 % higher C18:1 content than its donor parent. Genetic analysis showed that the two mutant lines were allelic, and the high C18:1trait exhibited a single gene segregation in F_3 families derived from the F_2 population between the mutant line 7488 and a wild-type line (Schierholt et al. 2000). Three AFLP markers were mapped near the locus of the *FAD2* gene in rapeseed genome (Schierholt et al. 2000). In a DH population derived from DMS100 (high in C18:1 content), a major QTL for C18:1 content was mapped on A5, accounting for 76.3 % of the total variation of the fatty acid. Furthermore, it was found that another minor QTL localized on the A1 accounted for 9.4 % of the variation (Hu et al. 2006).

Although different molecular markers linked to C18:1 and C18:3 were identified, the exact allelic variations for many of the markers are still unknown. It has been shown that in allopolyploid *B. napus* there are possibly four copies for *FAD2* and six for *FAD3* through RFLP analysis (Scheffler et al. 1997). However, the information on the sequences of these genes is incomplete. So far, only one *FAD2* mutation locus located on A5 was determined in *B. napus* (Hu et al. 2006). The development of allele-specific markers for C18:1 and C18:3 mutations is still a demanding task in *B. napus* due to its polyploidy nature,. In many cases, this type of markers has a low resolution, as multiple copies for the genes are available in *B. napus* genome. Therefore, it is important to identify all the copies in amphidiploid *B. napus* genome and to develop a more robust allele-specific type of markers for marker-assisted selection (MAS) in the breeding of high C18:1 and low C18:3.

The aims of this study were (1) to analyze the genetic control of high C18:1 and low C18:3 in a DH population derived from a cross between a spring-type cultivar with high C18:1 and low C18:3 contents and a winter-type cultivar of *B. napus*, (2) to better understand the gene structures of all the *FAD2* and *FAD3* copies in the *B. napus* genome, and (3) to develop allele-specific markers for high oleic and low linolenic contents. In this study, we identified a novel mutation in *FAD2* locus in the A-genome of *B. napus*. Based on the thorough annotation of *FAD2* and *FAD3* genes in three *Brassica* species, we developed highly reliable allele-specific markers for *FAD2* and single-nucleotide amplified polymorphisms (SNAP) markers for *FAD3* genes by utilizing the amplification refractory mutation system (ARMS) (Newton et al. 1989) and the SNAPER program (Drenkard et al. 2000). We demonstrated that the locus-specific markers are highly specific to distinguish different copies of the same gene and effective for the selection of high C18:1 and low

C18:3 in different breeding populations. The genome distribution information of the *FAD2* and *FAD3* genes will be valuable for breeding for C18:1 and C18:3 contents. Furthermore, successful applications of SNAP markers to *B. napus* will improve the efficiency of single nucleotide polymorphism (SNP) detection, which is widely available but often difficult to use with traditional PCR techniques in such a species with multiple loci of a same gene.

Materials and methods

Plant materials and phenotypic analysis

Two *B. napus* genotypes with distinct C18:1 and C18:3 contents were used in this study. One was SW Hickory, a spring-type *B. napus* variety with about 78 % of C18:1 and <3 % of C18:3 (a kind gift from Lantmännen SW Seed, Sweden), and another JA177 with about 64 % of C18:1 and 8 % of C18:3. A doubled haploid (DH) population including 238 DH lines was produced through microspore culture of F_1 buds of the cross SW Hickory × JA177. A random subset of 190 DH lines was sampled for the genetic analysis and mapping QTL of C18:1 and C18:3 contents. The population will henceforward be referred to as the SJ DH population (Fan et al. 2010).

The SJ DH population was grown during the winter-type oilseed rape growing season on the experimental farm of Huazhong Agriculture University, Wuhan, China for the three consecutive growing years (2006–2007, 2007–2008 and 2008–2009). Self-pollinated seeds were collected from individual DH plants in the year of 2006–2007, or from at least six plants in a DH line in the following seasons (2007–2008 and 2008–2009) for fatty acid profiling (for the convenience of description, the growing seasons are expressed as 2007–2009 thereafter). Fatty acid analysis was performed by using gas liquid chromatography (GC) with a Model 6890 GC analyzer (Agilent Technologies, Inc., Wilmington, DE, USA), following the protocol as described (Thies 1971). The content of individual fatty acid was expressed as the percentage of total fatty acids in mature seeds.

Molecular marker and QTL mapping

A linkage map covering the whole genome of *B. napus* has been constructed with the SJ DH population using SSR markers as described (Fan et al. 2010). The linkage map was used to localize the loci for C18:1 and C18:3 contents following the standard QTL mapping procedure (Fan et al. 2010). Significance thresholds at the 0.05 significance level were estimated on the basis of 1,000 permutations using the procedure as described (Churchill and Doerge 1994).

Bioinformatic analysis of *FAD2* and *FAD3* genes in *Brassica* species

To identify all putative *FAD2* and *FAD3* genes in *B. napus*, we searched for the sequences homologous to *Arabidopsis FAD2* (AT3G12120.1) and *FAD3* (AT2G29980.1) genes in a *Brassica* database (BRAD: http://brassicadb.org/brad/, Chinese Academy of Agricultural Sciences) containing *Brassica rapa* and *Brassica oleracea* genome sequences (Cheng et al. 2011). The search was conducted first with the BLASTn program using an *E* value $<1E{-}20$ and identity and query coverage >80 %. All hits from this search were then used to locate the homologous sequences in the NCBI nucleotide collections (http://www.ncbi.nlm.nih.gov) with the same parameters as above. The nonredundant sequences resulted from the hits were then collected and compared with known *FAD2* and *FAD3* genes in *Arabidopsis* .

The program FGENESH (HMM-based gene structure prediction tool) was used for gene prediction (http://www.softberry.com/). Sequence alignment was performed with CLUSTALX1.83 (Thompson et al. 1997) and phylogenetic analysis was applied to the alignment results. The similarity of the predicted protein sequences was compared based on a BLOSUM62 matrix (Henikoff and Henikoff 1993). Distance matrices were calculated based on Poisson correction, and trees were constructed with the neighbor-joining algorithm using the MEGA4.0 (Tamura et al. 2007). Bootstrap analysis with 1,000 replicates was performed to test the significance of nodes.

All gene nomenclature in this study followed the rules proposed by Østergaard and King (2008). Gene structure display server (Guo et al. 2007) was employed to generate the gene structure draws.

Cloning of *FAD2* and *FAD3* genes in *B. napus*

Genomic DNA was prepared from leaf tissues as described previously (Fan et al. 2010). For cDNA cloning, total RNA was extracted from the leaves of *B. napus* using TRIZOL (Invitrogen, Paisley, UK) and converted into first-strand cDNA following the manufacturer's instructions (TIANScript RT Kit, Beijing, China). The primers used for amplification of the genomic and/or cDNA fragments are listed in supplementary Table S1.

The PCR products were cloned into pMD18-T vector (Takara Corporation, Japan) according to the manufacturer's instructions. The M13F and M13R universal primers and the BigDye Terminator Cycle Sequencing v3.1 (Applied Biosystems, Foster City, CA, USA) were used for sequencing. Sequences were aligned using the computer program of SEQUENCHER 4.1.2 (Gene Codes Corporation, Ann Arbor, MI, USA).

SNAP primer design and allele-specific PCR

A SNAPER program (Drenkard et al. 2000) was used to design the SNAP primers. SNP among all possible *FAD3* genes from *B. napus* was used for the design of locus-specific primer, and the SNP of the same locus on two parents was used for the design of allele-specific primers. A locus-specific primer was then used as the forward primer together with an allele-specific primer as the reverse primer for a specific PCR reaction. The PCR products were separated by electrophoresis in a 1.0 % agarose gel stained with ethidium bromide for imaging.

Results

Genetic control of high oleic acid and low linolenic acid in the SJ DH population

Significant differences between the parents were detected for the contents of C18:1, C18:2 and C18:3 based on the data collected from three consecutive field seasons (2007–2009, Fig. 1). The distribution patterns of the three unsaturated C18 fatty acids from the three seasons were similar in the SJ DH population (Fig. 1), indicating that they were genetically stable across generations. The segregation patterns of C18:1 and C18:2 resembled each other. Specifically, the two fatty acids could be separated into two distinct groups at the cutoff of about 70.5 % for C18:1 and about 18.5 % for C18:2 (Fig. 1). The two groups in each of the two traits fit well into a 1:1 ratio across the three seasons, indicating the involvement of a major gene for the traits. A continuous distribution within the two groups (Fig. 1), on the other hand, suggested that there may be some minor genes involved in each trait and/or the contents of the two fatty acids were likely affected by the environment to some extents.

For the content of C18:3, it was difficult to classify individual DH lines into distinct groups, although the distribution curve showed two peaks in 2007 and 2009 seasons (Fig. 1). This distribution indicated that there was a more complicated genetic control for the C18:3 content in the population compared to the C18:1 content.

Genetic mapping of QTLs for the contents of three unsaturated C18 fatty acids

QTLs for the contents of C18:1, C18:2 and C18:3 in the SJ DH population were detected using the procedure of composite interval mapping as described in "Materials and methods". For C18:1, a major QTL with the largest effect, *OLEA5* on LG A5 (Fig. 2) was detected, which accounted for roughly 83 % of the total variation and the allele from

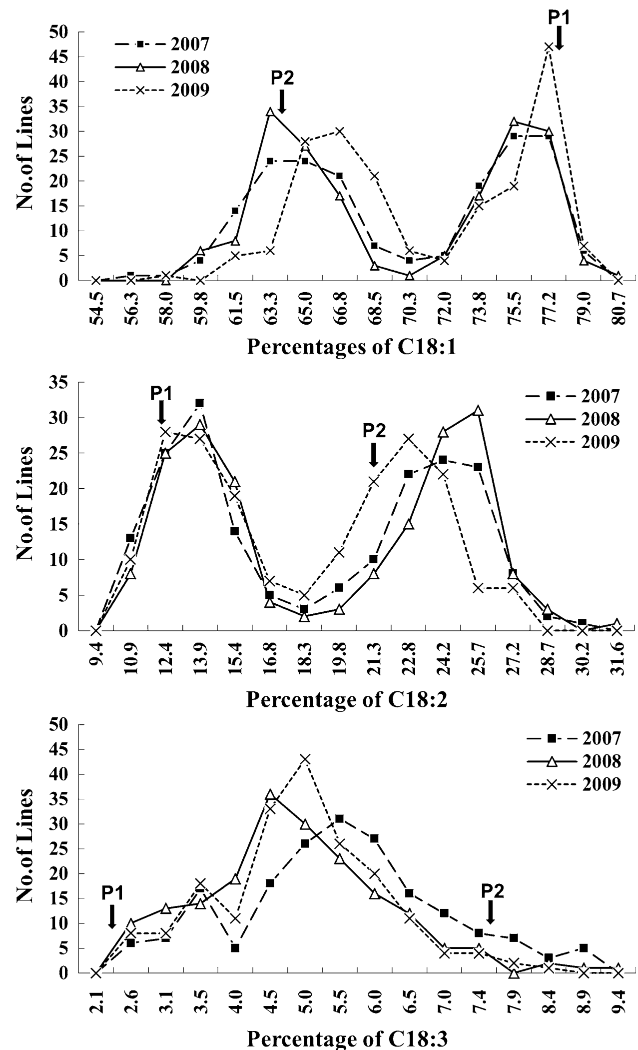

Fig. 1 Distribution of the three unsaturated C18 fatty acids measured in the SJ DH population derived from the cross of SW Hickory × JA177. The average measurement of the parents (P1, SW Hickory; P2, JA177) in 3 years was indicated by an *arrow* for each of the three fatty acids

SW Hickory increased C18:1 content by 5.4 % (additive effect) based on the 3-year average (Table 1). Another QTL for C18:1 was only detected in the year of 2009 and mapped on LG A3 (Fig. 2), which only showed a minor effect (2.5 %) to the phenotypic variation. Three QTLs for C18:2 content were resolved on A3, A5 and C4 in 3 years (Fig. 2; Table 1). Notably, LAA5, the QTL with the largest effect accounting for 88 % of the trait variation based on the 3-year average (Table 1), was located at the same position as the major QTL for C18:1 on LG A5 (Fig. 2). However, it had an opposite effect in which the allele from SW Hickory decreased C18:2 content by about 5.0 % (additive effect, Table 1). This suggested that the two alleles at this locus affected these two fatty acids differently. The other two QTLs identified in only one of the years showed small effects (<3 % contribution from each).

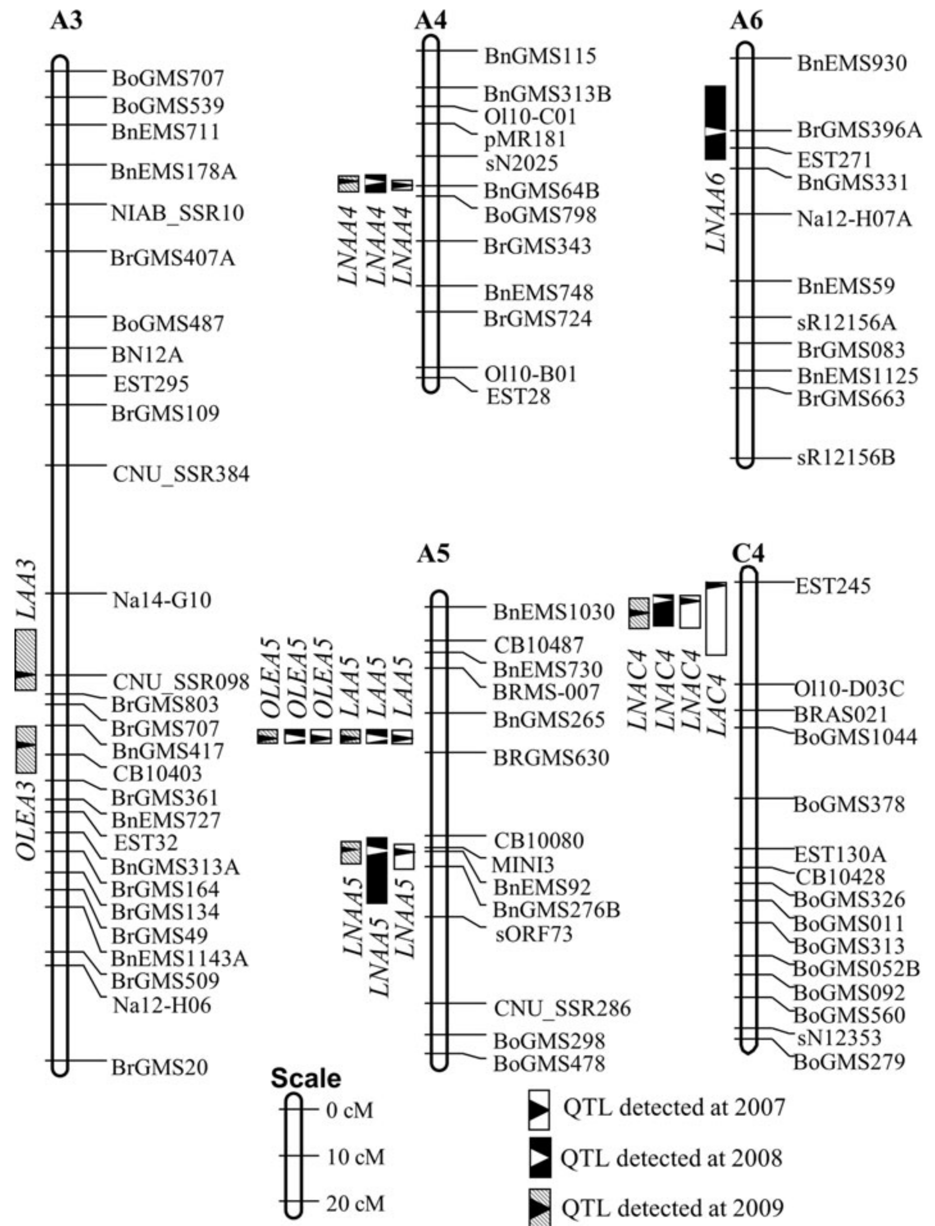

Fig. 2 The genetic linkage map and QTLs for the contents of C18:1, C18:2 and C18:3 identified in the SJ DH population in 3 years. The *bar* to the *left* of the LG indicates the 1-LOD confidence interval for the QTL and the *triangle* indicates the QTL peak position

Four QTLs for C18:3 content mapped on A4, A5, A6 and C4 collectively explained more than 90 % of the total phenotypic variation in each year (Fig. 2; Table 1). All of the alleles with reduced C18:3 content were contributed by SW Hickory. The largest effect was caused by a QTL (*LNAC4)* located closely to the *LAC4* with an additive effect of about 1 %, which explained over 60 % of the variation of C18:3 content based on the 3-year average. The QTL *LNAA4* on LG A4 also had a sizeable effect and accounted for 29 % of the variation, with the allele from SW Hickory reducing C18:3 content by 0.7 % in average. Jointly, the two QTLs accounted for over 90 % of the variation in C18:3 level in this population. The remaining two QTLs, *LNAA5* and *LNAA6*, were detected in 3 and 1 years, respectively, with their contributions ranging from 2.4 to 9.1 % of the phenotypic variation individually (Fig. 2; Table 1).

Identification of *FAD2* and *FAD3* genes in *Brassica* A- and C-genome

The conversions of oleic acid to linoleic acid and linoleic acid to linolenic acid were achieved through consecutive reduction reactions controlled by the fatty acid desaturase genes *FAD2* and *FAD3*, respectively (Arondel et al. 1992; Lemieux et al. 1990; Miquel 1992; Miquel et al. 1993; Okuley et al. 1994). Therefore, the high oleic content QTL, *OLEA5* may represent one of the copies of the *FAD2* gene and the low linolenic QTLs, *LNAA4* and *LNAC4*, two copies of the *FAD3* gene in *B. napus*. To gain a full

Table 1 QTLs for oleic acid (C18:1), linoleic acid (C18:2) and linolenic acid (C18:3) detected with SSR markers

Trait	QTL[a]	2007				2008				2009				3-year mean			
		CI and peak (cM)[b]	LOD	A[c]	R^2 (%)[d]	CI and peak (cM)	LOD	A	R^2 (%)	CI and peak (cM)	LOD	A	R^2 (%)	CI and peak (cM)	LOD	A	R^2 (%)
C18:1	*OLEA3*									142.0–151.5 (145.89)	3.2	0.9	2.5				
	OLEA5	27.0–29.8 (28.99)	71	5.6	79.9	27.0–29.6 (28.99)	81.7	5.7	83.6	27.0–29.6 (28.99)	71.7	4.9	83.1	27.0–29.5 (28.99)	89.5	5.4	82.6
C18:2	*LAA3*									121.3–133.5 (130.69)	3.7	−1.1	2.7				
	LAA5	27.0–29.6 (28.99)	81.1	−5.2	83.9	27.0–29.6 (28.99)	86.9	−5.3	86.9	27.0–29.6 (28.99)	67.2	−4.5	77.2	27.0–29.5 (28.99)	90.4	−5	87.9
	LAC4	0.0–14.7 (0.01)	3.7	0.9	2.2												
C18:3	*LNAA4*	27.9–29.9 (29.31)	23.6	−0.8	27.3	26.2–29.7 (29.11)	26.6	−0.7	27	27.1–30.3 (29.11)	21.2	−0.6	26.4	27.6–29.8 (29.31)	29	−0.7	28.8
	LNAA5	51.2–56.3 (53.09)	7.4	−0.4	7	49.9–64.5 (53.09)	11.1	−0.4	9.1	50.8–55.3 (53.09)	7.1	−0.3	7.2	52.0–86.7 (53.09)	11.1	−0.4	8.7
	LNAA6					7.0–22.0 (15.78)	3.3	−0.2	2.4								
	LNAC4	2.0–8.6 (4.01)	39.4	−1.2	57.7	2.0–8.3 (4.01)	46.4	−1.1	62.4	3.0–9.1 (6.01)	34.3	−0.9	57.4	3.6–8.7 (6.01)	51.4	−1	62.7

[a] QTL nomenclature is designated using the trait name initials followed by the LG number; an alphabetical letter a or b is added if more than one QTL are identified on one LG

[b] Confidence intervals were obtained by marking positions ±1 LOD from the peak (in bracket)

[c] A additive effect; positive additive effects indicate that the female parent (SW Hickory) allele increases the value of the trait

[d] Proportion of the phenotypic variation explained by the QTL

understanding of the *FAD2* and *FAD3* genes in *B. napus* genome, we searched for all possible copies of *FAD2* and *FAD3* gene in A-genome (*B. rapa*) and C-genome (*B. oleracea*) in databases, including a newly released genome sequence database for *B. rapa* (Cheng et al. 2011) and *B. oleracea* (http://brassicadb.org/brad/).

With *Arabidopsis FAD2* sequence (AT3G12120.1) as a query, each two *FAD2* gene loci was identified in *B. rapa* (A-genome) and *B. oleracea* (C-genome), respectively. All gene coding sequences (CDS) except for one (locus BGIScaffold000151 at A1 chromosome) from the search are predicted to encode a full open reading frame (ORF) with high sequence similarity to the *Arabidopsis FAD2* gene (Supplementary Fig. S1). The sequence of the putative *FAD2* locus located at the scaffold of BGIScaffold000151 in *B. rapa* genome contained a single-base deletion at the 164th base pair, a 2-bp insertion between the 188–189th base pairs, a 15-bp deletion between the 230–244th base pairs and a single-base deletion at the 424th base pairs, compared to other copies of putative *FAD2* genes (Supplementary Fig. S1), which resulted in a premature stop codon at the 411th position due to the frame shift starting at the 164th base pairs. It is likely that the corresponding sequence in BGIScaffold000151 may represent a pseudogene of *FAD2* in A-genome. These genes were named *BraA.FAD2.a* and *BraA.FAD2.b-like,* and *BolC.FAD2.a* and *BolC.FAD2.b*, respectively (Table 2), following the nomenclature rules proposed by Østergaard and King (2008). The predicted amino acids encoded by the genes shared high similarity (Supplementary Fig. S2).

Similarly, with the sequence of *Arabidopsis FAD3* gene (AT2G29980.1) as a query, each three *FAD3* gene loci were identified in *B. rapa* and *B. oleracea*, respectively. These genes were named *BraA.FAD3.a, BraA.FAD3.b, BraA.FAD3.c* and *BolC.FAD3.a, BolC.FAD3.b, BolC.FAD3.c,* respectively (Table 2). All of the identified *FAD3* loci were predicted to contain an intact CDS based on the sequence of the *Arabidopsis FAD3* gene (Supplementary Fig. S3; UTRs of the genes are not shown).

Based on the sequence and structure information of the *FAD2* genes in *B. rapa* and *B. oleracea*, the putative genomic and cDNA fragments of *FAD2* in *B. napus* were isolated through a homologous cloning strategy from JA177, the wild-type parent of the SJ DH population. Four complete genomic DNA sequences of *FAD2* were identified after PCR amplification, molecular cloning and sequencing (Supplementary Fig. S1). The sequences from cDNA cloned using the same pair of primers for each *FAD2* fragment were identical to corresponding genomic DNA sequences (data not shown), further demonstrating that *FAD2* is an intronless gene as reported in *Arabidopsis* (Okuley et al. 1994), *B. rapa* (Tanhuanpää et al. 1998) and *B. napus* (Hu et al. 2006). The cloned four *FAD2* nucleotides

Table 2 Genome distribution of *FAD2 and FAD3* in *B. napus, B. rapa* and *B. oleracea*

Gene identity[a]	Chr[b]	Sequences available	Accession	Source
BnaA.FAD2.a	A5[d]	gDNA and cDNA	JN992605[e]; JN992606	This study
BraA.FAD2.a	A5	gDNA and putative CDS	Bra034777	BRAD
BnaC.FAD2.a	C5[d]	gDNA and cDNA	JN992607	This study
BolC.FAD2.a	C5	gDNA and putative CDS	Bol010322	BRAD
BnaC.FAD2.b	C1[d]	gDNA and putative CDS	JN992609	This study
BolC.FAD2.b	C1	gDNA and putative CDS	Bol007285[c]	BRAD
BnaA.FAD2.b-like	A1[d]	gDNA	JN992608	This study
BraA.FAD2.b-like	A1	gDNA	BGIScaffold000151 no CDS in BRAD	BRAD
BnaC.FAD3.a	C4[d]	gDNA and cDNA	JN992617	This study
BolC.FAD3.a	C4	gDNA and putative CDS	Bol017645	BRAD
BraA.FAD3.a	A5	gDNA and putative CDS	Bra018348	BRAD
BnaA.FAD3.a	A5[d]	gDNA and cDNA	JN992616	This study
BnaC.FAD3.b	C4	gDNA and cDNA	JN992614; JN992615[e]	This study
BolC.FAD3.b	C4	gDNA and putative CDS	Bol033304	BRAD
BnaA.FAD3.b	A4	gDNA and cDNA	JN992612; JN992613[e]	This study
BraA.FAD3.b	A4	gDNA and putative CDS	Bra021599	BRAD
BolC.FAD3.c	C3	gDNA and putative CDS	Bol001628[c]	BRAD
BnaC.FAD3.c	C3[d]	gDNA and putative CDS	JN992611	This study
BraA.FAD3.c	A3	gDNA and putative CDS	Bra022767	BRAD
BnaA.FAD3.c	A3[d]	gDNA and putative CDS	JN992610	This study

[a] Gene nomenclature follows the rules proposed by Østergaard and King (2008)

[b] *Chr* chromosome. The chromosome locations of the *FAD2* and *FAD3* genes from *B. rapa* and *B. oleracea* are based on the sequencing information at *Brassica* database (BRAD, Cheng et al. 2011)

[c] Because the CDS of gene at the BRAD was incomplete, we annotated the CDS according to the genomic nucleotide sequence by using the gene prediction program FGENESH (http://linux1.softberry.com/)

[d] Chromosome location is deduced based on the synteny of *B. napus* and its two progenitor species, *B. rapa* and *B. oleracea*

[e] The DNA sequences cloned from the high oleic low linolenic parent SW Hichory

in *B. napus* were then named *BnaA.FAD2.a, BnaA.FAD2.b-like*, *BnaC.FAD2.a* and *BnaC.FAD2.b*, respectively (Table 2; Supplementary Fig. S1).

Six genomic DNA fragments corresponding to the six *FAD3* copies in A- and C-genome were isolated (Supplementary Fig. S3). The six copies of the *FAD3* gene in *B. napus* were designated *BnaA.FAD3.a, BnaA.FAD3.b, BnaA.FAD3.c, BnaC.FAD3.a, BnaC.FAD3.b* and *BnaC.FAD3.c*, respectively (Table 2; Supplementary Fig. S3). The genomic sequences of the six copies of *FAD3* gene in *B. napus* and their gene structure exhibited great variation from each other (Supplementary Fig. S3 and S4). However, at cDNA and protein levels, the sequences of the six *FAD3* genes and their encoding proteins shared high similarity with each other (Supplementary Fig. S3 and S5).

Phylogenetic analyses were performed based on the predicted protein sequences of *FAD2* and *FAD3* genes identified in this study (Table 2) to reveal evolutionary relationship among the copies. The three intact *FAD2* members from *B. napus* and their corresponding counterparts in two diploid species could be grouped into two clusters (*FAD2.a* and *FAD2.b*; Fig. 3a). Members belonging to each of the two categories were more distantly related than the members within a same category. For example, *BnaC.FAD2.a* and *BolC.FAD2.a* were more closely related to each other than members from the same species such as *BolC.FAD2.a* and *BolC.FAD2.b* (Fig. 3a). It seemed that the *FAD2.a* was more closely related to *Arabidopsis* *FAD2*.a (Fig. 3a). The protein members encoded by the *FAD3* gene family could be grouped into three categories (*FAD3.a*, *FAD3.b* and *FAD3.c*; Fig. 3b). Members belonging to the same category were more closely related to each other regardless of their genome and species origin (Fig. 3b); this is consistent with the tendency found in the *FAD2* gene family. The bootstrap values in the neighbor-joining tree for the members of *FAD3* from the C-genome were larger than the ones from the A-genome (Fig. 3b), indicating that the similarity of *BolC.FAD3* and *BnaC.FAD3* proteins is higher than *BraA.FAD3* and *BnaA.FAD3* proteins. Phylogenetic analysis with the cDNAs obtained very similar results (data not shown). From an evolutionary perspective, these results indicated

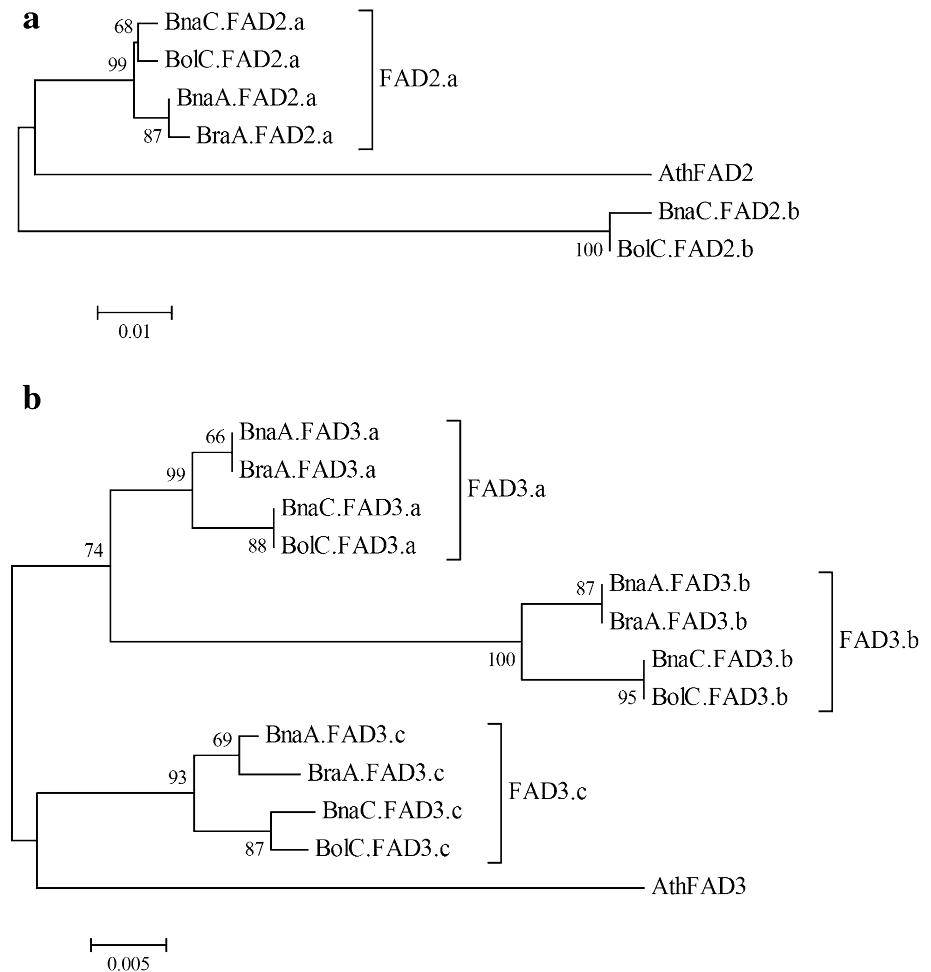

Fig. 3 Phylogenetic analyses of *FAD2* and *FAD3* genes in three *Brassica* species. Neighbor-joining tree is presented based on the deduced amino acid sequences of *FAD2* and *FAD3* genes in *B. rapa, B. oleracea* and *B. napus*. Bootstrap values (1,000 replications) are shown at each branch as a percentage. A branch length *scale bar* is shown beneath each tree

that the differentiation of more than one of *FAD2* and/or *FAD3* members might have occurred before the origin of two diploid species of *B. rapa* (AA) and *B. oleracea* (CC) as independent species, and this differentiation had been conserved during the evolution of *Brassica* species.

A novel mutation in the *BnaA.FAD2.a* gene responsible for the high oleic content in SW Hickory

With all the *FAD2* and *FAD3* gene sequences identified, experiments were performed to identify the polymorphism of *FAD2* gene in the two parents that shows a significant difference in the C18:1 content. For that purpose, sequences of 20–30 cloned *FAD2* fragments from two parents were compared. All the sequences except for *BnaA.FAD2.a* were the same (data not shown). Compared with the wild type, the sequence of *BnaA.FAD2.a* from SW Hickory contained a 4-bp insertion at the 567–568th base pair (Fig. 4). The difference was further confirmed with the cDNA comparison between the two parents (data not shown). The insertion resulted in a frame shift that led to a misreading and a premature termination of the translation at 667–669th (TGA) position (Fig. 4).

To verify that the mutated *BnaA.FAD2.a* in SW Hickory was indeed responsible for the high oleic trait, a SCAR marker was developed based on the difference between the two parents. Two primer pairs, YQ-*fad2a*-1 and YQ-*fad2a*-2, were designed to detect the deletion polymorphism in the SW Hichory and JA177 alleles of the *FAD2* gene (Table 3). The YQ-*fad2a*-1 forward primer contained the 4-bp insertion at its 3′-end to ensure the specificity to the SW Hichory allele. Similarly, the YQ-*fad2a*-2 forward primer included the 4-bp deletion to ensure the specificity to the JA177 allele. Together, the two pairs of primers could be used for the identification of all possible genotypes at this particular locus (Fig. 5a).

The SCAR marker was then used to map the mutated allele in the SJ DH population. The allele-specific marker of *FAD2* gene was co-located on the *OLEA5* with a higher LOD score (Supplementary Table S2; Supplementary Fig. S6), indicating that the mutation at *BnaA.FAD2.a* is responsible for the high oleic trait. In addition, rescanning of QTL for C18:1 by including the allele-specific marker also resulted in slight changes in QTL numbers for C18:1 (Supplementary Table S2; Supplementary Fig. S6).

```
P2         ATGTTAACGGTTCAGTTCACTCTCGGC----TGGCCTTTGTACTTAGCCTTCAACGTCTCGGGGAGACCTTACGACGGCGGCTTCGCTTG 626
            M  L  T  V  Q  F  T  L  G     W  P  L  Y  L  A  F  N  V  S  G  R  P  Y  D  G  G  F  A  C
P1         ATGTTAACGGTTCAGTTCACTCTCGGCAGCCTGGCCTTTGTACTTAGCCTTCAACGTCTCGGGGAGACCTTACGACGGCGGCTTCGCTTG 630
            M  L  T  V  Q  F  T  L  G  S  L  A  F  V  L  S  L  Q  R  L  G  E  T  L  R  R  R  L  R  L
DMS100     ATGTTAACGGTTTAGTTCACTCTCGGC----TGGCCTTTGTACTTAGCCTTCAACGTCTCGGGGAGACCTTACGACGGCGGCTTCGCTTG 484
            M  L  T  V  *  F  T  L  G     W  P  L  Y  L  A  F  N  V  S  G  R  P  Y  D  G  G  F  A  C
Consensus  ************ **************    ************************************************************ 482

P2         CCATTTCCACCCCAACGCTCCCATCTACAACGACCGTGAGCGTCTCCAGATATACATCTCCGACGCTGGCATCCTCGCCGTCTGCTACGG 716
             H  F  H  P  N  A  P  I  Y  N  D  R  E  R  L  Q  I  Y  I  S  D  A  G  I  L  A  V  C  Y  G
P1         CCATTTCCACCCCAACGCTCCCATCTACAACGACCGTGAGCGTCTCCAGATATACATCTCCGACGCTGGCATCCTCGCCGTCTGCTACGG 720
            P  F  P  P  Q  R  S  H  L  Q  R  P  *  A  S  P  D  I  H  L  R  R  W  H  P  R  R  L  L  R
DMS100     CCATTTCCACCCCAACGCTCCCATCTACAACGACCGTGAGCGTCTCCAGATATACATCTCCGACGCTGGCATCCTCGCCGTCTGCTACGG 574
             H  F  H  P  N  A  P  I  Y  N  D  R  E  R  L  Q  I  Y  I  S  D  A  G  I  L  A  V  C  Y  G
Consensus  ****************************************************************************************** 572
```

Fig. 4 Alignment of the partial genomic nucleotide sequences of the *BnaA.FAD2.a* gene cloned from SW Hickory (P1) and JA177 (P2). The nucleotide mutations between parents are highlighted in *black boxes*. The putative protein sequences are given below the nucleotide sequences; the *asterisk* denotes the termination codon

Table 3 Primer sequences of the molecular markers developed in the study

Primer pairs		Sequence (5′–3′)[a]	T_m (°C)	Size (bp)	
				P1	P2
YQ-*fad2a*-1	Forward	CAGTTCACTCTCGGCAGC	58	201	–
	Reverse	GCAACTCCTTGGACAGCA			
YQ-*fad2a*-2	Forward	CAGTTCACTCTCGGCTGG	58	–	258
	Reverse	GATCAAAACTAAGAACCCG			
YQ-*fad3a*-1	Forward	GGAGTTTCTCAGACATTCGC	54	797	–
	Reverse	GTGGTCGTCCAAGCCTCTC			
YQ-*fad3a*-2	Forward	GGAGTTTCTCAGACATTCGT	51	–	796
	Reverse	GTGGTCGTCCAAGCCTCTC			
YQ-*fad3c*-1	Forward	CCTTGGTACAGAGGCAACA	52	428	–
	Reverse	TTGCCACCAAACTTTCCAGT			
YQ-*fad3c*-2	Forward	CCTTGGTACAGAGGCAATG	55	–	428
	Reverse	TTGCCACCAAACTTTCCAGT			

Size PCR product size, *P1* SW Hickory, *P2* JA177, "–" no amplification

[a] The additional mismatch bases in primer pairs were underlined

Previously, Hu et al. (2006) identified a single base pair conversion (C to T) in *FAD2* copy located on LG5. In our study, however, this nucleotide in the two parents exhibited no such difference (Fig. 4). The mutation in SW Hickory, thus, represents a previously uncharacterized allele mutation for the *FAD2* gene in *Brassica* genome.

Identification of the mutated *FAD3* alleles for low linolenic acid in SW Hickory

Since two major QTLs responsible for low linolenic acid had been located at LG A4 and LG C4 (Fig. 2; Table 1), comparison of the sequence difference between the two parents were focused on the loci localized on the LGs where the mapped two QTLs reside. Based on the distributions of the corresponding locus in A- and C-genome, we searched for sequence differences at the loci of *BnaA.FAD3.b* and *BnaC.FAD3.b* between two parents.

Two single-nucleotide mutations between the parents SW Hickory and JA177 were identified in the locus of *BnaA.FAD3.b*. The first one was a T-to-C substitution in exon 2 leading to a synonymous mutation, and the second a C-to-T transition in exon 7, which resulted in an amino acid substitution from arginine to cysteine (Fig. 6a). The later mutation (C to T) was the same as the one identified by Mikolajczyk et al. (2010) in the winter-type low C18:3 mutant DH219/05, and the differences were further confirmed by comparing the cDNAs from the two parents (data not shown). At the genomic level, there was a 10-base insertion in the 6th intron of SW Hichory in addition to several single-nucleotide alterations between the two parents (data not shown).

Sequence analysis and alignment revealed only a single-nucleotide mutation between SW Hickory and JA177 for another target sequence, *BnaC.FAD3.b*, which was a G-to-A substitution at the 3048th site of genomic DNA in SW Hickory. Further, comparison with the cDNA sequences from the two parents revealed that the mutation position in genomic DNA corresponded to the 5′ splice site of the sixth intron of *BnaC.FAD3.b* (Fig. 6b). The mutation, thus,

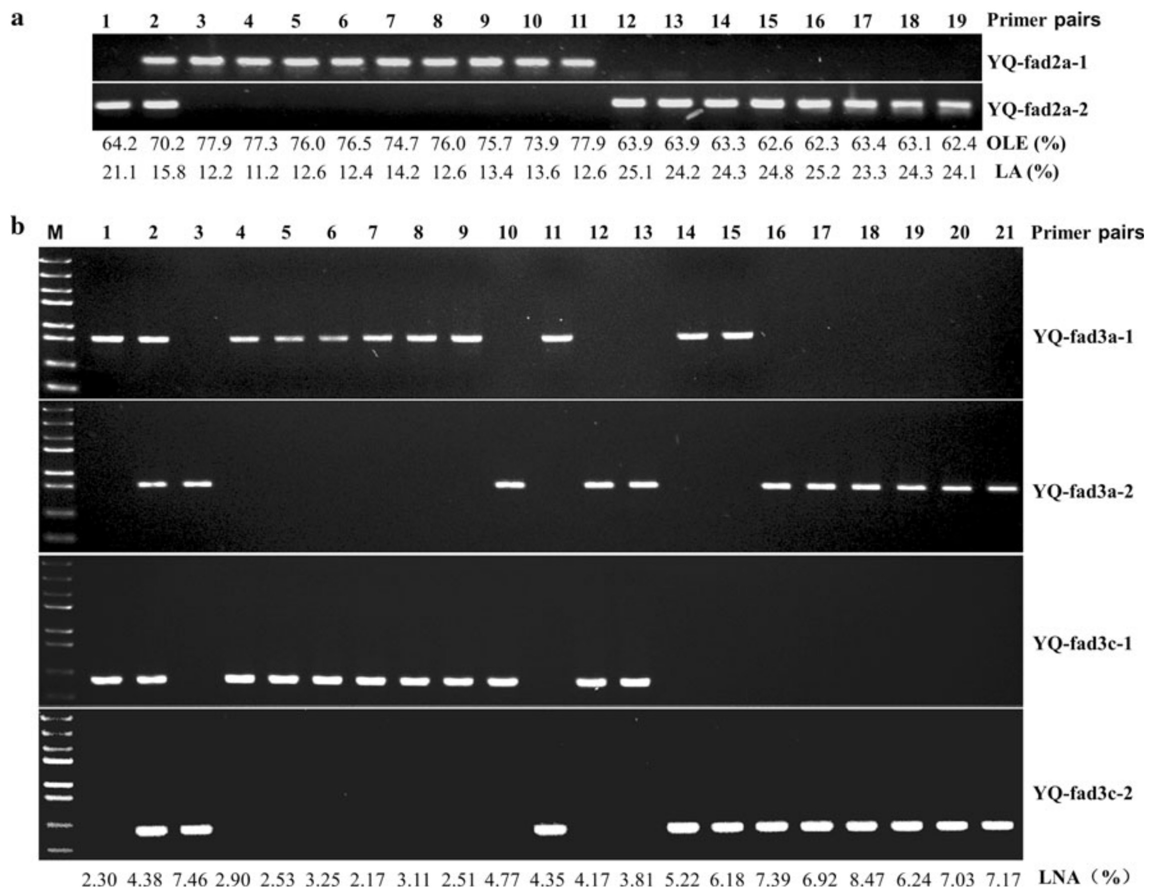

Fig. 5 PCR products amplified from the allele-specific markers for the *fad2* and *fad3* genes developed in the study. The PCR products were separated by electrophoresis in a 1.5 % agarose gel and stained with ethidium bromide. **a** PCR products amplified from primer pairs YQ-*fad2*a-1 and YQ-*fad2*a-2; lane 1, JA177; lane 2, the hybrid (F_1) from the cross of SW Hickory × JA177; lane 3, SW Hickory; lanes 4–19, DH lines 59, 23, 61, 65, 23, 67, 46, 76, 167, 74, 47, 62, 121, 55, 14 and 51 from SJ DH population; the average oleic and linoleic acid contents of the corresponding lines in 3 years were listed below. **b** PCR products amplified from primer pairs YQ-*fad3a*-1, YQ-*fad3a*-2, YQ-*fad3c*-1 and YQ-*fad3c*-2; *M* DNA ladder, from top to bottom: 5,000, 3,000, 2,000, 1,500, 1,000, 800, 500 and 300 bp; lane 1, SW Hickory; lane 2, the hybrid (F_1) from the cross of SW Hickory × JA177; lane 3, JA177; lanes 4–21, DH lines 1, 2, 4, 6, 21, 30, 8, 3, 23, 33, 37, 38, 17, 18, 20, 22, 39 and 40 from SJ DH population; the average linolenic acid contents of the corresponding lines in 3 years were listed below

resulted in an abnormal splicing with the entire sixth intron being retained in the mature *BnaC.FAD3.b* transcript (Fig. 6b), which was the same as the mutations identified by Hu et al.(2006) and Mikolajczyk et al. (2010). Another copy of the *FAD3* gene on C4, *BnaC.FAD3.a* did not show polymorphisms at both genomic and cDNA levels between the two parents (data not shown).

Development of SNAP markers of mutated *FAD3* gene for low linolenic content

The SNPs present in the two copies of *fad3* genes were subsequently used for he development of their allele-specific markers. However, our initial efforts using primers containing only a single base pair difference often resulted in a low resolution in PCR reactions, making it difficult to detect the differences among the parents and the progeny lines (data not shown). To overcome this problem, a procedure called SNAP (Drenkard et al. 2000) was introduced. In this procedure, additional mismatch bases (2–3) were introduced at bases in the immediate front of the SNP site to reduce the non-specific amplification. In the case of *BnaA.FAD3.b*, the T/C polymorphism in the exon 2 together with two mismatch bases were introduced into the 3′-end of forward primers, YQ-*fad3a*-1 and YQ-*fad3a*-2, respectively (Table 3). We found that high sequence similarity between the two copies of *FAD3* exists near their polymorphic sites (Supplementary Fig. S3). To achieve locus specificity, we made use of the 1626th G/T polymorphism between the two copies of *FAD3* and introduced additional mismatched base pairs in the reverse primers (Supplementary Fig. S3) in addition to the mismatches in the respective forward SNAP primers (Table 3). By doing so, both locus and allele specificities were successfully obtained (Fig. 5b).

Fig. 6 Alignment of the partial genomic nucleotide sequences of *BnaA.FAD3.b* (**a**) and *BnaC.FAD3.b* (**b**) genes cloned from SW Hickory (P1) and JA177 (P2). The cDNA sequences are shown in *capital letters*; the introns are in *lowercase letters*. The nucleotide mutations between parents are highlighted in *black boxes*. The putative protein sequences are given below the nucleotide sequences; the *asterisk* denotes the termination codon

Using the same approach, two primer pairs based on the 3048th G/A polymorphism in *BnaC.fad3.b* gene were developed, in which YQ-*fad3c*-1 and YQ-*fad3c*-2 primer pairs targeted the SW Hickory and JA177 alleles of *BnaC.FAD3.b* gene, respectively. Furthermore, we used the 3439th G/A polymorphism between *BnaA.FAD3.b* and *BnaC.fad3.b* (Supplementary Fig. S3) to design the locus-specific primers (Table 3). These allele-specific markers for the two *FAD3* loci amplified distinct bands from SW Hickory, JA177 and their hybrid with expected band size and specificity (Fig. 5b; Table 3). Although each of two primer pairs for a particular locus above was dominant, the joint use of the two primer pairs for each locus was able to operate in a codominant manner (Fig. 5b).

The codominant markers of *BnaA.FAD3.b* and *BnaC.FAD3.b* genes were subsequently mapped to the respective linkage groups in the SJ DH population, and were found to be located exactly on the *LNAA4* and *LNAC4* QTL peaks, respectively (Supplementary Fig. S6). Rescanning of QTLs for C18:3 including these allele-specific markers resulted in higher LOD scores of the *LNAA4* and *LNAC4* in each of the three seasons (Supplementary Table S2), suggesting that *BnaA.FAD3.b* and *BnaC.FAD3.b* are likely to be the genes underlying the two QTLs of *LNAA4* and *LNAC4*. In addition, rescanning of QTL for C18:3 by including the allele-specific markers also resulted in some slight changes in QTL effects for C18:3 (Supplementary Table S2; Supplementary Fig. S6).

Validation of the allele-specific markers for C18:1 and C18:3 contents

With the SCAR marker for *BnaA.FAD2.a*, the 190 DH lines from the SJ DH population could be classified into two types of genotype: homozygote for the SW Hickory allele (*aa*) and for the JA177 allele (*AA*). Individuals with *aa* genotype exhibited significantly higher C18:1 content and lower C18:2 content, while there was no significant difference in C18:3 (Table 4). The average C18:3 content in *AA* and *aa* groups was lower than the wild type because some low C18:3 individuals were included in each group (Table 4).

Similarly, with the SNAP markers for *BnaA.FAD3.b* and *BnaC.FAD3.b*, the 190 DH lines could be grouped into four types of homozygous genotype. C18:3 content in the groups consisting of the individuals with a mutant allele at one of the two loci were significantly lower than the wild type (Table 4), confirming the genetic effects from the mutation of *BnaA.FAD3.b* and/or *BnaC.FAD3.b* gene. As expected, the individuals with only one mutated locus, either *BnaA.FAD3.b* or *BnaC.FAD3.b,* had significant higher C18:3 content than the ones with both mutated loci (Table 4, data not shown). Interestingly, the mutation in *BnaC.FAD3.b* resulted in a more obvious reduction of C18:3 than *BnaA.FAD3.b* (the difference was significant at $P = 0.01$ level, data not shown). Apparently, with these allele-specific markers, C18:3 level can be manipulated

Table 4 Percentages (mean ± SD) of three unsaturated C18 fatty acids in groups identified with three allele-specific markers in the SJ DH population

Trait	Year	*BnaA.FAD2.a* (AA)		*BnaA.FAD3.b* (BB) and *BnaC.FAD3.b* (CC)					AA and BB and CC	
		AA[a]	aa	BBCC	bbcc	bbCC	BBcc	BBcc and bbCC	AABBCC	aabbcc
		97[b]	93	39	33	76	42	118	16	13
C18:1	2007	63.6 ± 2.5	74.7 ± 2.0**	69.59 ± 5.62	68.45 ± 6.18	68.82 ± 6.30	69.31 ± 5.90	68.99 ± 6.14	63.42 ± 2.21	75.29 ± 1.57**
	2008	63.4 ± 2.0	75.0 ± 1.9**	69.50 ± 6.47	67.99 ± 5.65	68.74 ± 6.25	69.86 ± 5.98	69.13 ± 6.15	62.18 ± 2.18	74.73 ± 2.31**
	2009	65.5 ± 2.2	75.3 ± 1.9**	70.76 ± 5.64	70.23 ± 5.17	69.91 ± 5.38	70.65 ± 5.21	70.18 ± 5.31	64.74 ± 2.95	76.09 ± 1.24**
C18:2	2007	23.3 ± 2.2	12.8 ± 1.9**	15.93 ± 5.03	20.93 ± 5.63**	18.05 ± 5.6*	18.29 ± 5.52*	18.13 ± 5.55*	21.47 ± 2.10	14.36 ± 1.26**
	2008	23.9 ± 1.9	13.1 ± 1.6**	16.59 ± 5.59	21.47 ± 5.24**	18.77 ± 5.54*	18.10 ± 5.66	18.54 ± 5.57*	22.92 ± 2.04	15.22 ± 1.82**
	2009	22.1 ± 2.1	13.0 ± 1.8**	15.67 ± 4.79	19.49 ± 4.72**	17.83 ± 4.89*	17.54 ± 4.84*	17.72 ± 4.86**	20.78 ± 2.50	14.14 ± 1.18**
C18:3	2007	5.4 ± 1.2	5.1 ± 1.4	6.93 ± 1.16	3.21 ± 0.66**	5.43 ± 0.82**	4.91 ± 0.75**	5.25 ± 0.83**	7.42 ± 1.10	2.93 ± 0.67**
	2008	4.8 ± 1.4	4.5 ± 1.2	6.17 ± 1.21	2.83 ± 0.37**	4.81 ± 0.64**	4.17 ± 0.59**	4.59 ± 0.69**	6.75 ± 1.31	2.55 ± 0.37**
	2009	4.8 ± 1.1	4.5 ± 1.1	5.94 ± 1.05	3.06 ± 0.53**	4.88 ± 0.61**	4.43 ± 0.56**	4.72 ± 0.63**	6.42 ± 1.10	2.69 ± 0.38**

* and ** indicate a significant difference at the 0.05 probability level and 0.01 probability level between wild type (JA177) and corresponding mutated alleles based on *t* test, respectively

[a] Letters (A, B, C) mean the genotypes with the homozygous alleles from JA177 (capital letters) and SW Hickory (small letters), respectively

[b] Sample size

with great flexibility (for example, from 3 to 5 %). In all cases, the C18:1 content among the groups was not significantly different, further demonstrating the independent identification of the target locus with the markers for *FAD2* and *FAD3* genes.

Discussion

Whole genome annotation for the *FAD2* and *FAD3* genes in three *Brassica* species

Through bioinformatic analysis and molecular cloning, we identified for the first time all predicted copies of *FAD2* and *FAD3* gene in amphidiploid species *B. napus* as well as in two diploid *Brassica* species, *B. rapa* and *B. oleracea* (Table 2). The annotation of the two important gene families in these three species will be valuable for the understanding of fatty acid synthesis control. For example, the relative contribution from different alleles to the magnitude of oleic or linolenic content and the specific roles for each different copies could be compared through the analysis of gene expression with the aid of the sequence information. Through data mining and molecular cloning, we showed that there is a naturally mutated ORF for one of the *FAD2* locus in A-genome of both *B. rapa* and *B. napus*. The locus contained five insertions or deletions (Supplementary Fig. S1), which resulted in a frame shift starting at the 164th base pairs and consequently a likely non-functional gene/protein. The effect of such a mutation on the *FAD2* control in *Brassica* species awaits further investigation. A thorough understanding of the functional difference of the alleles will also facilitate the genetic manipulation of the important traits of C18:1 and/or C18:3 content through genomic techniques such as the targeting induced local lesion in genome (TILLING, McCallum et al. 2000). The sequence information will also be useful for the development of molecular markers for breeding.

The sequences of the *FAD2* gene and encoding amino acids in three examined *Brassica* species were highly similar among the loci (Supplementary Fig. S1 and S2). On the other hand, the genomic sequences of the *FAD3* gene are rather divergent (Supplementary Fig. S3). This pattern is consistent with two diploid species and *B. napus,* indicating that the differentiation of the sequences had already occurred before the origin of two diploid species, *B. rapa* and *B. oleracea*. Interestingly, at protein levels, both FAD2 and FAD3 members among the loci share high similarities (Supplementary Fig. S2 and S5), suggesting that plants have evolved to produce a mechanism to maintain the stability at protein levels of the two gene families. It is likely that the conserved protein structures are necessary for plants to retain functional proteins to survive natural selection.

This feature could be utilized in genetic engineering of fatty acid by the application of antisense or RNA interfere (Rania) techniques to manipulate the content of oleic and linolenic acids. On the other hand, whether the different copies of the *FAD2* and *FAD3* genes are differentially expressed in different plant tissues awaits further studies, which could provide a more complete function understanding of the gene families.

A novel mutation responsible for the high oleic acid in SW Hickory

Although molecular markers have been previously linked to the variation of oleic content in *B. napus* (Hu et al. 2006; Javidfar et al. 2006; Schierholt et al. 2000), only one allele of *FAD2* gene with determined sequence mutation was reported prior to the current study (Hu et al. 2006). A single-nucleotide substitution (C to T) in DMS100 was identified, which could result in a premature termination of the *FAD2* peptide chain and a truncated polypeptide with only 185 amino acids instead of a wild-type polypeptide with 384 amino acids (Hu et al. 2006). We present here a novel allele mutation of the gene different in nature from the previous one. In our study, a 4-bp insertion at 567–568th base pair was identified in SW Hickory, a spring high C18:1 cultivar. This insertion led to a misreading and premature termination of *BnaA.FAD2.a* (Fig. 4). The mutation, thus, results in an obvious aberrant transcript for the mutated locus as shown with cDNA cloning analysis, which in turn can be utilized to develop the allele-specific marker (Figs. 4, 5a). These two different types of mutation in the same allele resulted in a similar increase of oleic content (about 78 %), supporting the notion that the *BnaA.FAD2.a* encodes a functional fatty acid reductase in *B. napus*. Our genotyping analysis with the allele-specific marker shows a high correlation of the marker genotype and phenotypic performance (Table 4), providing further evidence that the mutation in *BnaA.FAD2.a* identified in this study is responsible for the high C18:1 content in SW Hickory. It is known that *FAD2* catalyzes the desaturation reaction of oleate that results in the formation of a new double bond at the 12th position and thus of linoleic acid (Broun et al. 1998; Covello and Reed 1996). Blocking of the desaturation reaction by interference of *FAD2* function eventually leads to more accumulation of oleic acid and a reduced linoleic acid in the fatty acid synthesis pool.

The four copies of the *FAD2* gene in *B. napus* exhibited high similarity in both cDNA and protein levels (Supplementary Fig. S1 and S2). The fact that a single locus mutation in *BnaA.FAD2.a* led to a significant increase of C18:1 content in seeds highlights the importance of the gene and its encoded protein. An interesting but still unresolved question is why the other two copies could not completely compensate for the failure of *BnaA.FAD2.a* mutation. So far, all identified markers have been linked to the locus on chromosome 5 in the A-genome (Hu et al. 2006; Javidfar et al. 2006; Scheffler et al. 1997; Schierholt et al. 2000; Tanhuanpää et al. 1998; Teutonico and Osborn 1994; Uzunova et al. 1995). The relationship of this locus with other *FAD2* loci would be an interesting topic for further characterization of *FAD2* gene family in *Brassica* genome. A better understanding of the gene nature will also facilitate the further enhancement of C18:1 level in *B. napus*, which could be beneficial for the application of rapeseed oil as a raw material for biodiesel.

Utilization of SNAP markers in amphidiploid *B. napus*

SNP markers are the powerful tools used to identify the allelic variation in model plants like *Arabidopsis* and crops for positional cloning and MAS (Morgante and Salamini 2003). A common strategy for distinguishing an allele SNP is to use allele-specific primers designed to include a 3′-terminal nucleotide that corresponds with the site of the SNP (Ugozzoli and Wallace 1992). Such an allele-specific primer, thus, matches exactly with the target allele and has a 3′-terminal mismatch with the non-specific allele. However, a single base pair difference at the 3′-terminus in an allele-specific primer is often not sufficient for a robust discrimination between the two SNP alleles, even in *Arabidopsis* with a relatively small genome (Drenkard et al. 2000). The low amplification efficiency in the traditional PCR analysis is primarily caused by the different mismatches at the 3′-end that could result in different extension efficiencies by DNA polymerase under normal PCR conditions (Kwok et al. 1990; Li et al. 1990; Newton et al. 1989). This would eventually lead to a low resolution for allele specific detection.

In a genome with complicated structure, like amphidiploid *B. napus*, the application of allele-specific markers based on SNPs becomes even more difficult. In many cases, a single base pair change at the 3′-end of the non-specific allele is not sufficient to create a reliable discrimination between the two alleles (Cha et al. 1992; Drenkard et al. 2000; Kwok et al. 1990, 1994). There could be four to six copies for a single gene in *B. napus*, as demonstrated in this study by the *FAD2* and *FAD3* genes. The copies of the genes usually share high sequence similarity in *B. napus.* Our initial efforts to use primers with a single 3′-terminal nucleotide match to the SNP at *BnaA.FAD3.b* and *BnaC.FAD3.b* resulted in a low resolution to distinguish the two alleles, although there existed several SNPs between high and low linolenic parents in the two *FAD3* copies (data not shown). Previously, Hu et al. (2007) used a pair of primers specific to *fad3c* (corresponding to *BnaC.FAD3.b* in this study) with only a single base pair mismatch for the

RT-PCR analysis in *B. napus*, but non-specific amplifications were observed, which was proposed to be likely due to the amplifications of other *FAD3* copies in A-genome (Hu et al. 2007). By using SNAP markers, we were able to overcome the non-specific amplification in resolving the SNPs in *Bna.FAD3* genes (Fig. 5b). To the best of our knowledge, this is the first report of the successful use of such a type of SNP marker in *B. napus*. We are now routinely using the strategy for our mapping work in *B. napus*. Over the last decades, different types of molecular markers have been developed for C18:1 and C18:3 variations (Hu et al. 1995, 1999, 2006; Javidfar et al. 2006; Jourdren et al. 1996b; Mikolajczyk et al. 2010; Rajcan et al. 1999; Schierholt et al. 2000; Tanhuanpää et al. 1995, 1996, 1998; Tanhuanpää and Schulman 2002; Thormann et al. 1996) in *Brassica* species. Given the complexity of the *B. napus* genome, markers with high specificity and high throughput natures are most desirable. In this regard, SNP markers offer greater advantages compared to other types of molecular markers. In this study, we developed SNAP markers which allowed for the routing application of a locus-specific primer together with an allele-specific primer through 3′-mismatch technique (Table 3). Such a combination made it possible to utilize the SNP markers reliably in a PCR-Gel system and significantly reduced the interference of other copies of the homologous gene in amphidiploid *B. napus*. As such, we are able to detect heterozygous statuses in *FAD3* loci based on a single base pair difference. It is plausible that the SNAP marker technique is also applicable with other genes in diploid and amphidiploid *Brassica* species.

Acknowledgments The authors thank Mr. Bo Gertsson, Lantmännen SW Seed, for providing the seeds of SW Hickory. We are grateful to Miss Christy Zhou at Duke University for critical reading of the manuscript. This research was financially supported by China National High-tech R&D Program (2006AA101A113) and the earmarked fund for Modern Agro-industry Technology Research System (nycytx-00503) to YZ.

References

Arondel V, Lemieux B, Hwang I, Gibson S, Goodman HM, Somerville CR (1992) Map-based cloning of a gene controlling omega-3 fatty acid desaturation in *Arabidopsis*. Science 258:1353–1355

Barret P, Delourme R, Brunel D, Jourdren C, Horvais R, Renard M (1999) Low linolenic acid level in rapeseed can be easily assessed through the detection of two single base substitution in *fad3* genes. In: Proceedings of the 10th international rapeseed congress: new horizons for an old crop. Canberra, Australia. Available at http://www.regional.org.au/au/gcirc/4/385.htm

Broun P, Shanklin J, Whittle E, Somerville C (1998) Catalytic plasticity of fatty acid modification enzymes underlying chemical diversity of plant lipids. Science 282:1315–1317

Cha RS, Zarbl H, Keohavong P, Thilly WG (1992) Mismatch amplification mutation assay (MAMA): application to the c-H-ras gene. PCR Methods Appl 2:14–20

Cheng F, Liu S, Wu J, Fang L, Sun S, Liu B, Li P, Hua W, Wang X (2011) BRAD, the genetics and genomics database for *Brassica* plants. BMC Plant Biol 11:136

Churchill GA, Doerge RW (1994) Empirical threshold values for quantitative trait mapping. Genetics 138:963–971

Covello PS, Reed DW (1996) Functional expression of the extraplastidial *Arabidopsis thaliana* oleate desaturase gene (*FAD2*) in *Saccharomyces cerevisiae*. Plant Physiol 111:223–226

Drenkard E, Richter B, Rozen S, Stutius L, Angell N, Mindrinos M, Cho R, Oefner P, Davis R, Ausubel F (2000) A simple procedure for the analysis of single nucleotide polymorphisms facilitates map-based cloning in *Arabidopsis*. Plant Physiol 124:1483–1492

Fan C, Cai G, Qin J, Li Q, Yang M, Wu J, Fu T, Liu K, Zhou Y (2010) Mapping of quantitative trait loci and development of allele-specific markers for seed weight in *Brassica napus*. Theor Appl Genet 121:1289–1301

Guo A, Zhu Q, Chen X, Luo J (2007) GSDS: a gene structure display server. Yi Chuan 29:1023–1026

Henikoff S, Henikoff JG (1993) Performance evaluation of amino acid substitution matrices. Proteins 17:49–61

Hu J, Quiros C, Arus P, Strass D, Robbelen G (1995) Mapping of a gene determining linolenic acid concentration in rapeseed with DNA-based markers. Theor Appl Genet 90:258–262

Hu J, Li G, Struss D, Quiros C (1999) SCAR and RAPD markers associated with 18 carbon fatty acids in rapeseed, *Brassica napus*. Plant Breed 118:145–150

Hu XY, Sullivan-Gilbert M, Gupta M, Thompson SA (2006) Mapping of the loci controlling oleic and linolenic acid contents and development of *fad2* and *fad3* allele-specific markers in canola (*Brassica napus* L.). Theor Appl Genet 113:497–507

Hu XY, Sullivan-Gilbert M, Gupta M, Thompson SA (2007) G-to-A mutation at a 5′ splice site of fad3c caused impaired splicing in a low linolenic mutant of canola (*Brassica napus* L.). Plant Biotechnol 24:397–400

Javidfar F, Ripley VL, Roslinsky V, Zeinali H, Abdmishani C (2006) Identification of molecular markers associated with oleic and linolenic acid in spring oilseed rape (*Brassica napus*). Plant Breed 125:65–71

Jourdren C, Barret P, Brunel D, Delourme R, Renard M (1996a) Specific molecular marker of the genes controlling linolenic acid content in rapeseed. Theor Appl Genet 93:512–518

Jourdren C, Barret P, Horvais R, Delourme R, Renard M (1996b) Identification of RAPD markers linked to linolenic acid genes in rapeseed. Euphytica 90:351–357

Kwok S, Kellogg D, McKinney N, Spasic D, Goda L, Levenson C, Sninsky J (1990) Effects of primer-template mismatches on the polymerase chain reaction: human immunodeficiency virus type 1 model studies. Nucleic Acids Res 18:999–1005

Kwok S, Chang S-Y, Sninsky JJ, Wang A (1994) A guide to the design and use of mismatched and degenerate primers. PCR Methods Appl 3:S39–S47

Lemieux B, Miquel M, Somerville C, Browse J (1990) Mutants of *Arabidopsis* with alterations in seed lipid fatty acid composition. Theor Appl Genet 80:234–240

Li H, Cui X, Arnheim N (1990) Direct electrophoretic detection of the allelic state of single DNA molecules in human sperm by using the polymerase chain reaction. Proc Natl Acad Sci USA 87:4580–4584

Matthäus B (2006) Utilization of high-oleic rapeseed oil for deep-fat frying of French fries compared to other commonly used edible oils. Eur J Lipid Sci Technol 108:200–211

McCallum CM, Comai L, Greene EA, Henikoff S (2000) Targeted screening for induced mutations. Nat Biotechnol 18:455–457

Mikolajczyk K, Dabert M, Karlowski W, Spasibionek S, Nowakowska J, Cegielska-Taras T, Bartkowiak-Broda I (2010) Allele-specific SNP markers for the new low linolenic mutant genotype of winter oilseed rape. Plant Breed 129:502–507

Miquel M (1992) *Arabidopsis* mutants deficient in polyunsaturated fatty acid synthesis. Biochemical and genetic characterization of a plant oleoyl-phosphatidylcholine desaturase. J Biol Chem 267:1502–1509

Miquel M, James D, Dooner H, Browse J (1993) *Arabidopsis* requires polyunsaturated lipids for low temperature survival. Proc Natl Acad Sci USA 90:6208–6212

Morgante M, Salamini F (2003) From plant genomics to breeding practice. Curr Opin Biotechnol 14:214–219

Newton C, Graham A, Heptinstall L, Powell S, Summers C, Kalsheker N, Smith J, Markham A (1989) Analysis of any point mutation in DNA. The amplification refractory mutation system (ARMS). Nucleic Acids Res 17:2503–2516

Okuley J, Lightner J, Feldmann K, Yadav N, Lark E (1994) *Arabidopsis FAD2* gene encodes the enzyme that is essential for polyunsaturated lipid synthesis. Plant Cell 6:147–158

Østergaard L, King GJ (2008) Standardized gene nomenclature for the *Brassica* genus. Plant Methods 4:10

Pinzi S, Garcia I, Lopez-Gimenez F, Luque de Castro M, Dorado G, Dorado M (2009) The ideal vegetable oil-based biodiesel composition: a review of social, economical and technical implications. Energy Fuels 23:2325–2341

Rajcan I, Kasha K, Kott L, Beversdorf W (1999) Detection of molecular markers associated with linolenic and erucic acid levels in spring rapeseed (*Brassica napus* L.). Euphytica 105:173–181

Rakow G (1973) Selektion auf linol- und linolen-saureghalt in rapssamen nach mutagener behand-lung. Z Planzen 69:205–209

Röbbelen G, Nitsch A (1975) Genetical and physiological investigations on mutants for polyenoic fatty acids in rapeseed, *Brassica napus* L. I. Selection and description of new mutants. Z Pflanzenzucht 75:92–105

Scarth RT, Tang JH (2006) Modification of oil using conventional and transgenic approaches. Crop Sci 46:1225–1236

Scarth R, McVetty P, Rimmer S, Stefansson B (1988) Stellar low linolenic-high linoleic acid summer rape. Can J Plant Sci 68:509–511

Scheffler J, Sharpe A, Schmidt H, Sperling P, Parkin I, Lühs W, Lydiate D, Heinz E (1997) Desaturase multigene families of *Brassica napus* arose through genome duplication. Theor Appl Genet 94:583–591

Schierholt A, Becker H, Ecke W (2000) Mapping a high oleic acid mutation in winter oilseed rape (*Brassica napus* L.). Theor Appl Genet 101:897–901

Tamura K, Dudley J, Nei M, Kumar S (2007) MEGA4: molecular evolutionary genetics analysis (MEGA) software version 4.0. Mol Biol Evol 24:1596–1599

Tanhuanpää P, Schulman A (2002) Mapping of genes affecting linolenic acid content in *Brassica rapa* ssp. *oleifera*. Mol Breed 10:51–62

Tanhuanpää P, Vilkki J, Vilkki H (1995) Association of a RAPD marker with linolenic acid concentration in the seed oil of rapeseed (*Brassica napus* L.). Genome 38:414–416

Tanhuanpää P, Vilkki J, Vilkki H (1996) Mapping of a QTL for oleic acid concentration in spring turnip rape (*Brassica rapa* ssp. *oleifera*). Theor Appl Genet 92:952–956

Tanhuanpää P, Vilkki J, Vihinen M (1998) Mapping and cloning of *FAD2* gene to develop allele-specific PCR for oleic acid in spring turnip rape (*Brassica rapa* ssp. *oleifera*). Mol Breed 4:543–550

Teutonico R, Osborn T (1994) Mapping of RFLP and qualitative trait loci in *Brassica rapa* and comparison to the linkage maps of *B. napus, B. oleracea*, and *Arabidopsis thaliana*. Theor Appl Genet 89:885–894

Thies W (1971) Schnelle und einfache Analysen der Fettsäurezusammensetzung in einzelnen Rapskotyledonen I. Gaschromatographische und papierchromatographische Methoden. Z Pflanzenzücht 65:181–202

Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, Higgins DG (1997) The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic Acids Res 25:4876–4882

Thormann C, Romero J, Mantet J, Osborn T (1996) Mapping loci controlling the concentrations of erucic and linolenic acids in seed oil of *Brassica napus* L. Theor Appl Genet 93:282–286

Ugozzoli L, Wallace RB (1992) Application of an allele-specific polymerase chain reaction to the direct determination of ABO blood group genotypes. Genomics 12:670–674

Uzunova M, Ecke W, Weissleder K, Röbbelen G (1995) Mapping the genome of rapeseed (*Brassica napus* L.). I. Construction of an RFLP linkage map and localization of QTLs for seed glucosinolate content. Theor Appl Genet 90:194–204

Warner K, Knowlton S (1997) Frying quality and oxidative stability of high-oleic corn oils. J Am Oil Chem Soc 74:1317–1322

Wittkop B, Snowdon R, Friedt W (2009) Status and perspectives of breeding for enhanced yield and quality of oilseed crops for Europe. Euphytica 170:131–140