Mol Breeding (2019) 39: 75
https://doi.org/10.1007/s11032-019-0971-x

Promoter polymorphism in *FAE1.1* and *FAE1.2* genes associated with erucic acid content in *Brassica juncea*

Navinder Saini · Yashpal · Murali Krishna Koramutla · Naveen Singh · Satbeer Singh · Rajendra Singh · Sangita Yadav · Ramcharan Bhattacharya · Sujata Vasudev · D. K. Yadava

Received: 30 June 2018 / Accepted: 2 April 2019 / Published online: 11 May 2019

Abstract *FAE1*, the gene encoding fatty acid elongase, has been implicated to play a decisive role in the synthesis of erucic acid. Allelic variations in this gene, such as single nucleotide polymorphisms (SNPs), insertions/deletions (InDels), and insertions of transposable element (TE), are predominantly responsible for variability in erucic acid content in the seed, among the Brassica genotypes. *Brassica juncea* being an amphidiploid, contains two paralogs, *FAE1.1* and *FAE1.2*. Although the coding DNA sequences (CDS) of *FAE* gene is highly conserved (> 95% identical), the promoter region is found to be polymorphic. The sequence variability in the promoter region of *FAE1* across different *Brassica* spp. suggests their likely involvement in the regulation of the gene expression. This study was aimed to evaluate polymorphism in the upstream region of *FAE1.1* and *FAE1*.2 genes, across the low erucic acid (LEA) and high erucic acid (HEA) cultivars, and to develop molecular markers based on this variability. The upstream regions of *FAE* were sequenced from LEA cultivar "Pusa Mustard 30" and HEA cultivar "Pusa Bold." A 28-bp deletion in the promoter of *FAE1.1* and a 340-bp insertion of a transposon-like element in the *FAE1.2* gene promoter were discovered in LEA genotype. Markers based on the sequence variability in the promoter regions of *FAE1.1* and *FAE1.2* were found to completely co-segregate with the seed erucic acid content. These markers can be effectively used in marker-assisted selection for development of low erucic acid cultivars in *B. juncea.*

Keywords *Brassica juncea* · *FAE1* · Erucic acid · Promoter · MAS · Molecular marker

Electronic supplementary material The online version of this article (https://doi.org/10.1007/s11032-019-0971-x) contains supplementary material, which is available to authorized users.

N. Saini (✉) · Yashpal · N. Singh · S. Singh · R. Singh · S. Vasudev · D. K. Yadava
Division of Genetics, ICAR- Indian Agricultural Research Institute (IARI), Pusa Campus, New Delhi 110012, India
e-mail: navin12@gmail.com

M. K. Koramutla · R. Bhattacharya
ICAR-National Research Centre on Plant Biotechnology, New Delhi, India

S. Yadav · D. K. Yadava
Division of Seed Science and Technology, ICAR-IARI, New Delhi, India

Introduction

Rapeseed-mustard constitutes a major source of vegetable oil globally. The seed oil of *Brassica* spp. possesses a very low amount of saturated fatty acid (SAFA) and a balanced proportion of polyunsaturated fatty acid (PUFA) as compared to other edible oils. However, the presence of high concentration (35–50%) of erucic acid (C22:1) in mustard oil renders it undesirable for human nutrition, as it causes myocardial infarction and increased blood cholesterol (Renard and Mcgregor 1992; Mortuza et al. 2006). Therefore, deriving plant types with low erucic acid in seed oil is a long-standing breeding objective in this crop. "Canola" type oil with

low erucic acid and low glucosinolate content, introduced in Canada (Mag 1983; Abbadi and Leckband 2011) and other regions, has been proposed as a safe mustard oil for human consumption, and the oil cake fit for animal feed. The reduced erucic acid content in total fatty acid profile increases the oleic acid (C18:1) fraction, as the nonfunctional fatty acid elongase 1 (*FAE1*) genes halt the conversion of C18:1 to C20:1 or C22:1. This, in turn, results in a better ratio of linoleic and linolenic fatty acids in the oil (Jagannath et al. 2011). In India, an attempt is being made to reduce the erucic acid (~ 50%) fraction, to a level of the internationally acceptable standard (< 2%), in the conventional genotypes of *Brassica juncea*. The "zero erucic mustard" (ZEM) was the first genotype identified for low erucic acid content in *B. juncea* (Kirk and Oram 1978). Its derivatives are being used now in breeding programs for development of LEA cultivars through phenotypic selection, followed by tedious biochemical analysis.

The erucic acid content in *Brassica* is governed by *FAE1* gene. It encodes for the enzyme β-ketoacyl-CoA synthase (KCS), which catalyzes the first of the four-step enzymatic reactions towards the synthesis of very long-chain monounsaturated fatty acids (VLCMFAs), in erucic acid biosynthesis pathway (James et al. 1995; Millar and Kunst 1997; Gupta et al. 2004). The diploid *Brassica* species (*B. rapa* and *B. nigra*) have one copy of *FAE1* gene, while the amphidiploid species (*B. napus* and *B. juncea*) have two copies with additive effect (Lühs et al. 1999; Gupta et al. 2004; Yan et al. 2015). Changes in the coding sequences (CDS) or regulatory regions, resulting from SNPs, InDels, and insertion of TE, are primarily responsible for the loss of function of *FAE1* gene, consequently reducing the VLCMFAs in seeds (Roscoe et al. 2001; Katavic et al. 2002; Chiron et al. 2015). In *B. napus*, a 4-bp deletion in the *FAE1* CDS leads to frameshift mutation causing a premature stop after the 466th amino acid residue (Wu et al. 2008). A reverse mutation in *FAE1* gene, through site-directed mutagenesis, converting back phenylalanine to a serine residue, at position 282, restored the elongase activity and erucic acid formation. This suggests that Ser 282 is a highly conserved amino acid in β-ketoacyl-CoA synthase enzyme, and in turn is required for the synthesis of erucic acid (Katavic et al. 2002). *FAE 1* CDS region is highly conserved (1521 bp) across species, with the exception of a few SNPs that have been reported to distinguish *FAE1.1* and *FAE1.2* in LEA and HEA genotypes of *B. juncea* (Gupta et al. 2004). These SNPs have been successfully used in marker-assisted breeding of *B. juncea*, after converting into CAPS markers (Saini et al. 2016).

However, the sequence comparison of the promoter region of *FAE1* of different *Brassica* species (Zeng and Cheng 2014; Yan et al. 2015; Chiron et al. 2015) provides insight into the differences in regulation of genes controlling the erucic acid content in the seed. In *B. rapa*, a deletion of 28 bases approximately 1300-bp upstream, the *FAE1* start codon co-segregates with the LEA trait in the segregating population, whereas in *Sinapis alba*, the LEA phenotype is attributed to the insertion of a transposable element and epigenetic modification in the promoter region, which affect the functionality of *FAE1* gene (Zeng and Cheng 2014; Fukai et al. 2019).

SNPs and InDels in gene or genome are evolving as a useful tool for marker-assisted breeding (Garcés-Claver et al. 2007). Although the CAPS marker developed for *FAE1* gene was successfully used in the breeding program (Saini et al. 2016), but when compared to simple PCR–based markers, CAPS markers are labor and cost intensive which restrict their routine use by the breeders. In this study, we have developed simple PCR–based markers by exploiting the variation in the upstream region of the *FAE1* genes. These markers are useful in marker-assisted selection of plants to improve oil quality in *B. juncea*.

Materials and methods

Materials and fatty acid profiling

A diverse collection comprising of 27 genotypes of *B. juncea* (AABB), *B. napus* (BBCC), *B. rapa* (AA), and *B. nigra* (BB) were taken for this study (Table 1). A backcross population (BC_1F_2) developed by crossing LEA variety Pusa Mustard 30 (PM 30) with Pusa Bold (Table 2) was used for validation and association studies. Gas chromatography (GC) by Perkin Elmer Clarus 600 with flame ionization detector (FID) was used for analyzing seed erucic acid content of parental genotypes and backcross population. During the analysis, the injector and detector temperatures were kept at 250 °C, while column temperature was programmed at a gradient of 10 °C increase per minute from an initial temperature of 150 °C until it reached 270 °C and maintained thereafter. Peaks of methyl esters of the fatty acid were

Table 1 Genotyping and phenotyping for *FAE1.1* and *FAE1.2* genes and erucic acid content in different genotypes

Species	Erucic acid content	Genotype	Varieties
B. juncea	High (30–43%)	$E_1E_1E_2E_2$	Pusa Vijay, Pusa Bold, Pusa Agrani, NRCDR-02, DRMRIJ-31, Laxmi, RH0749
	Intermediate (~20%)	$E_1E_1e_2e_2$/ $e_1e_1E_2E_2$	LS-1, Bio-YSR, Donskaja, BEC144
	Low (< 2%)	$e_1e_1e_2e_2$	PM 24, PM 30, PM 31, RLC-1, RLC-2, RLC-3, PDZ 4, PDZ 5, Pusa Karishma, PM 29, Heera
B. nigra	30–33%	E_2E_2	NG1, NG2, NG3
B. rapa	25–30%	E_1E_1	Tobin1, Rapa dwarf1
	0.5%	e_1e_1	Tobin2

identified by comparing with the known standards run under similar separation conditions (Sujata et al. 2008).

Analysis of molecular marker

DNA extraction was done from young expanding leaves of each individual genotype following the standard CTAB method (Doyle and Doyle 1990). DNA quantification was done using nanodrop™ (Nanodrop Technologies, USA) and quality was assessed by electrophoresis. DNA was diluted to a final concentration of 20 ng/μl with nuclease-free water and stored at − 20 °C. PCR was carried out in 10 μl reaction cocktail containing 25 ng of genomic DNA, 1 unit of Taq DNA polymerase (Vivantis, Malaysia.), 1× PCR assay buffer with 1.5 mM $MgCl_2$, 10 pmol of each primer (Forward and Reverse), and 100 μM of dNTPs mix (Thermo Scientific); and finally, the volume was made up using sterile distilled water. An ABI thermocycler was used for PCR cycling. The PCR cycles consisted of an initial denaturation step of 5 min at 94 °C, followed by 35 cycles of 30s denaturation at 94 °C, 30s annealing at 55 °C, and 45s extension at 72 °C and the final extension at 72 °C for 7 min. The amplified products were separated by electrophoresis on 3.5% metaphor agarose gel in 1× TBE buffer (Tris-borate 89 mM, boric acid 89 mM, EDTA 2 mM, pH 8.0) for 4 h at 80 V. A 100-bp DNA ladder (MBI Fermentas, USA) was used as size standard, to determine the size of the fragments. DNA staining was done with ethidium bromide and visualized using a gel documentation system. Genotyping was done with CAPS markers (Saini et al. 2016).

Table 2 Genotyping and phenotyping of BC_1F_2 population derived from PM 30 × Pusa Bold for erucic acid trait.

Parents and BC_1F_2 plants	Genotype	Phenotype (% erucic acid)
Pusa Bold	$E_1E_1E_2E_2$	43.10
PM 30	$e_1e_1e_2e_2$	0.21
29	$e_1e_1e_2e_2$	< 2.0
38	$E_1e_1e_2e_2$	6–14
31	$e_1e_1E_2e_2$	4–11
24	$E_1e_1E_2e_2$	18–32

Re-sequencing of *FAE1* upstream region and development of markers

The upstream sequences of *FAE1.1* gene of "A" genome of *B. rapa* (KP718763.1, KF999623.2, KF999615.2, F999632.2) and *B. napus* (XM_013798715.1) were downloaded from NCBI GenBank. These sequences were then compared using Multiple Sequence Alignment by CLUSTALW (http://www.genome.jp/tools-bin/clustalw) and NCBI blast program. For re-sequencing in selected genotypes, viz., Pusa Bold (HEA 43.1% erucic acid) and PM 30 (LEA 0.21% erucic acid), PrimerQuest Tool (https://eu.idtdna.com/PrimerQuest) was used to design primers (Table 3). The PCR-amplified products of both *B. juncea* (AABB genome) genotypes were sequenced. Similarly, the upstream sequence of *FAE1.2* gene on chromosome 7B of *B. juncea* (LFQT01001323.1) and *B. nigra* (LFLV01000642.1) was taken for *in silico* analysis and resequenced.

RNA isolation and cDNA synthesis

Total RNA from the flower and developing seed (immature embryos about 10 days after flowering and mature embryos about 25 days after flowering) of *B. juncea*

Table 3 Primer sequences used in the study for PCR and qRT-PCR analysis

Primer name	Sequence	Amplicon size
FAE1.1P	F-CGGAAGACTTGCTCTTCTTA	HEA–409 bp
	R-GTTATTTGGTGCATATAATACATA	LEA–381 bp
FAE1.2P	F-TCAGATCCCAGAAGACTTGTTAAT	HEA–534 bp
	R-ATGTATATATGTGTATGAATGTAGTACTCG F-GATCTGGATGTTCATAAGG	LEA–399 bp
Primer for RT		
GAPDH	F-CTAACTGCCTTGCTCCACTT R-TGTCTTCTGAGTTGCAGTGATAG	101 bp
FAE1-RT	F-TAACCATCGCTCCACTCTTTG R-TCAAGAAGTCAAGCCACGAC	219 bp

varieties PM 30 and Pusa Bold was isolated using TRI reagent (MRC, USA) according to the manufacturer's instructions. The residual DNA contamination in isolated RNA samples was eliminated by treating with RNase-free DNase (Qiagen, USA) as per the manufacturer's instructions. Quality and quantity of the RNA were determined by spectrophotometer (ND-1000 Nanodrop Technologies, USA). RNA samples having OD value greater than 2.0 at 260/230 were used for further analysis. RNA integrity was verified by resolving the samples on the agarose gel (1.8% in 1× TAE) through electrophoresis. The first strand cDNA was synthesized from 2 μg of total RNA using a cDNA synthesis kit (Takara Bio Inc., Japan) according to the manufacturer's instruction.

Quantitative real-time RT-PCR

To perform qRT-PCR reactions, SYBR green detection chemistry was deployed in a real-time PCR machine (Applied Biosystems, USA). KAPA SYBR® FAST qPCR Master Mix (2×) Kit (KAPA Biosystems, USA) was used to perform qRT-PCR. In brief, 20 μl of the reaction mixture was constituted with 2 μl diluted cDNA, 10 μl qPCR Master Mix (2×), 0.4 μl of ROX high reference dye (50×), and 0.4 μl of 10 μM of forward and reverse primers each. PCR cycling was carried out at an initial denaturation for 3 m at 95 °C, followed by 40 cycles consisting of three steps as 95 °C for 10 s, 60 °C for 30 s, and 72 °C for 30 s. Immediately, melt curve analysis of PCR products was carried out by a constant increase of temperature between 60 and 95 °C. Glyceraldehyde-3-phosphate-dehydrogenase (*GAPDH*) gene transcript was used in qRT-PCR normalization. All qRT-PCR experiments were carried out using three biological replicates with three technical replicates each time.

Results

In silico analysis of *FAE1* gene

The coding sequence (CDS) region and proteins of homologs and orthologs of *FAE1* genes of *Brassica* species including *B. rapa*, *B. nigra*, *B. oleracea*, *S. alba*, *Crambe hispanica*, *B. elongate*, *C. filiformis*, *Eruca vesicaria*, *B. tournefortii*, *B. napus*, and *B. juncea* were downloaded from NCBI GenBank and aligned by CLUSTALW. The result of the alignment showed that the CDS region is highly conserved having more than 94% nucleotide identity, while protein homology was > 95% across the *Brassica* species. The CDS of *FAE1* spanning over 1521 bp codes for 507 amino acids long polypeptide.

Analysis of *FAE1* upstream sequences

The putative promoter region of *FAE1* was found to be highly variable among *Brassica* spp. The size of the upstream promoter region varied from 1.0 to 1.3 kb. The first 600 bp of the promoter region, immediately upstream to the CDS region, was found to be conserved among the diploid species *B. rapa*, *B. nigra*, *B. oleracea*, and *S. alba*, whereas the region further upstream to this conserved segment was highly variable having plenty of

InDels. The promoter sequence of *FAE1.1* and *FAE1.2* gene (Fig. 1) was about 98% similar to their diploid progenitors, viz., *B. rapa* and *B. nigra*, respectively.

Analysis of FAE1.1 promoter

The upstream region of FAE1.1 gene was sequenced using by newly designed primers corresponding to the conserved sequences in the promoter region of the *FAE1.1* gene of *B. rapa* (NCBI GenBank accession nos. KP718763.1, KF999623.2, KF999615.2, KF999632.2) (Supplementary Fig. 1). The sequence information from *B. juncea* varieties "Pusa Mustard 30" and "Pusa Bold" exhibited deletion of "AT-" rich regions of *FAE1.1* promoter in PM 30. Moreover, the same InDels, as reported earlier in *B. rapa* (Yan et al. 2015) were also present in the *B. juncea*. In addition, several new SNPs, not reported earlier in *B. rapa* by Yan et al. (2015), were also detected in *B. juncea* genotypes. On the basis of polymorphism, one primer set *FAE1.1P* (Fig. 2 and Table 3) was found to be polymorphic between LEA (381 bp) and HEA (409 bp) genotypes. This primer set, however, did not amplify the allele in *B. nigra*, indicating that it is specific to genome "A."

Analysis of FAE1.2 promoter

The upstream region of *FAE1.2* gene on chromosome 7B of *B. juncea* (LFQT01001323.1) and *B. nigra* (LFLV01000642.1) was used for in silico analysis. The 1172-bp region upstream of CDS of *FAE1.2* was identified as a putative promoter sequence. Primers were designed and the upstream promoter region was PCR amplified and sequenced using the DNA of "Pusa Mustard 30" and "Pusa Bold" varieties. Sequence alignment of the amplified upstream regions revealed a 340-bp InDel between "Pusa Mustard 30" and "Pusa Bold" (Supplementary Fig. 2). The 340-bp insertion in Pusa Mustard 30 composed of a miniature TA-inverted repeat forming a stem-loop. It also showed "TA and TAA" sequences at the insertion site, which is the typical structure of a MITE. The sequence of insertion when analyzed by BLAST against Plant MITE database (http://pmite.hzau.edu.cn) revealed homology (6e−41) with super-family Tc1/Mariner transposon of *B. rapa*. Initially, primer sets were designed to develop a PCR-based marker system to utilize this polymorphism in the promoter region of *FAE1.2* gene. The primer set *FAE1.2 P* amplified an amplicon of 534 bp in HEA and 852 bp in LEA genotypes. Although there was a clear polymorphism between high and low genotypes, this primer set failed to generate expected amplicons in the F_1 plants generated by crossing HEA and LEA genotypes. To our surprise, when genomic DNAs of HEA and LEA genotypes were mixed together in vitro, only HEA allele was amplified, although both the amplicons could be amplified independently with DNA from either HEA or LEA genotypes. For circumventing this uncertainty, a number of diverse primer sets were tried under variable stringency of PCR conditions but none of them could amplify both the alleles from heterozygous individuals. The evanescing nature of LEA allele could be attributed to cross-hybridization of template DNAs from two different genotypes during PCR which formed hetero-duplex thus hindering its amplification. Michu et al. (2010) reported hetero-duplex formation during mixed-template polymerase chain reaction (PCR) using universal primers may cause serious problems in several PCR–based analyses. We attempted to solve the problem by developing an intervention of the hetero-duplex formation. Therefore, in addition, to forward and reverse primers, an extra forward primer corresponding to the

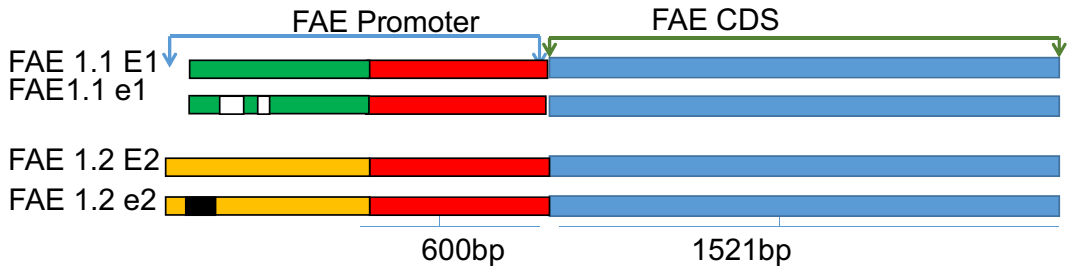

Fig. 1 Schematic representation of the promoter and CDS of the *FAE 1* paralogs. Blue boxes represent the CDS and red boxes indicate the similar sequences in the promoter region of *FAE1* paralogs. Green and yellow boxes represent the polymorphism in the *FAE1* promoter region; white boxes indicate the deleted regions in *FAE1.1* and black box represents the TE insertion in the *FAE1.2* promoter region

nucleotide sequence of the transposon was designed. Thus, the use of three primers (two forward corresponding to each allele and a common reverse primer) led to the amplification of a 399-bp amplicon instead of 852 bp in LEA genotypes (Fig. 2). Furthermore, the concentration of the reverse primer was increased from 10 to 15 pmol in the reaction cocktail. An appropriate combination of both of these approaches led to the simultaneous amplification of both the alleles.

Expression profile of *FAE1* genes

The relative transcript level of *FAE* genes in different reproductive tissues of *B. juncea* varieties grouped under low and high erucic acid types was estimated based on real-time qRT-PCR using gene-specific primers (Fig. 3). Since the CDS of *FAE1.1* and *FAE1.2* genes are highly conserved without any InDels and only a few SNPs differentiate both alleles, hence their transcripts cannot be detected separately, so only one primer set was used for transcript analysis of both genes simultaneously (Table 3). The level of transcript in each sample was represented as a multifold ratio to the level of transcript in flower tissue (Fig. 3). In the case of Pusa Bold, the maximum level of *FAE* transcript was observed to increase multifold in matured embryos when compared to the transcript level in flower; however, no significant difference in the *FAE* transcript levels was observed between the samples of the immature embryo and the flower tissue (Fig. 3a). In PM 30, the transcript level of *FAE* in immature and mature embryos was drastically downregulated, compared to the transcript level in flower (Fig. 3b). Thus, the relative transcript level of *FAE* corroborated the relative level of erucic acid content in oil of the LEA and HEA varieties of *B. juncea* in the mature embryos (Fig. 3).

Validation of polymorphism based on sequence variability in *FAE* upstream region of *B. juncea*

Twenty-seven genotypes, including commercial cultivars, improved elite lines, exotic germplasm, and genotypes of diploid species *B. nigra* and *B. rapa*, were phenotyped for erucic acid content in oil (Table 1). These accessions were also genotyped using CAPS markers developed by Saini et al. (2016). The biochemical phenotypes were perfectly corresponding to the genotypes revealed by the CAPS markers in all the accessions taken in this study. These genotypes were also examined for the allelic variation due to structural differences in the upstream promoter region of *FAE1.1* and *FAE1.2*. The markers based on promoter polymorphism distinctly differentiated the genotypes between LEA and HEA group. As expected, the *FAE1.1* promoter–based marker could not get amplified in *B. nigra* while *FAE 1.2* promoter–based marker could not be amplified both in *B. rapa* and in *B. napus*, confirming the genome-specific nature of these markers.

Genotyping using promoter-based markers was also performed in order to test the recombination and the reliability of these markers using the backcross population (BC_1F_2) developed by crossing LEA variety "Pusa Mustard 30" with HEA variety "Pusa Bold" (Table 2). The markers perfectly distinguished the parents and co-segregated with the biochemical phenotype in the backcross progeny, showing no recombination.

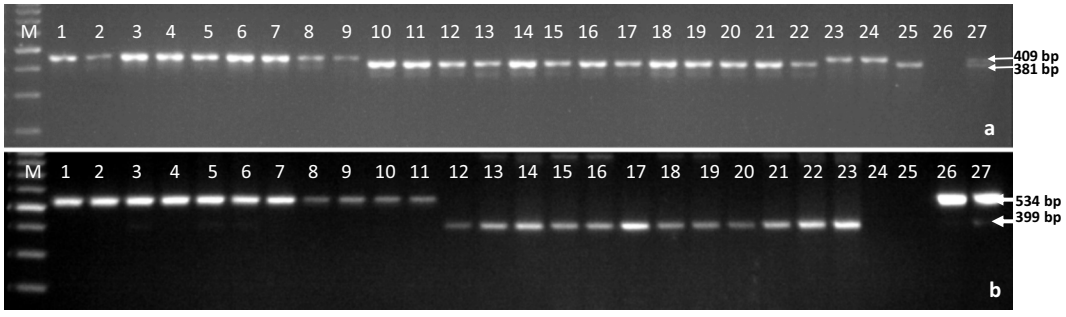

Fig. 2 Amplification pattern of *FAE1*.1 (**a**) and *FAE1*.2 (**b**) promoter markers in *Brassica* genotypes namely 1–Pusa Vijay, 2–Pusa Bold, 3–Pusa Agarani, 4–NRCDR-02, 5–DRMRIJ-31, 6–Laxmi, 7–RH0749, 8–Pusa Jagannath, 9–Bio-YSR, 10–Donskaja, 11–BEC-144, 12–PM 24, 13–PM 30, 14–PM 31, 15–RLC-1, 16–RLC-2, 17–RLC-3, 18–PDZ 4, 19–PDZ 5, 20–Pusa Karishma, 21–PM 29, 22–Heera, 23–LS-1, 24–Tobin1, 25–Tobin2, 26–NG3, and 27–F_1(Pusa Bold × PM 30)

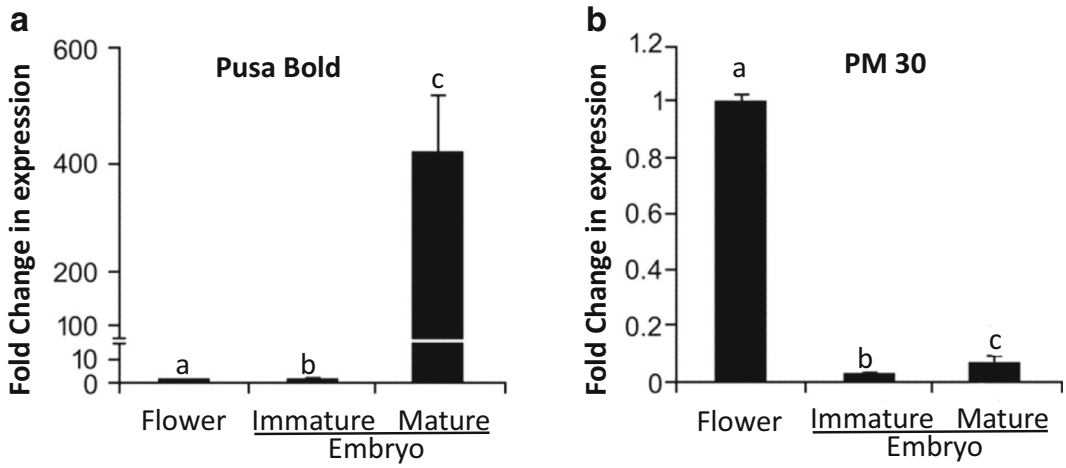

Fig. 3 The expression analysis of *FAE1* gene in flower and embryo tissues of HEA and LEA genotypes of *B. juncea*. Total RNA was isolated from the flower, immature, and mature embryos of Pusa Bold (**a**) and PM-30 (**b**) and assayed for the *FAE* expression by qRT-PCR using *GAPDH* as a normalizer. Fold change values represent mean ± SE ($3n = 9$) and comparison of means was carried out by Student's *t* test ($p < 0.05$). Different letters indicate significantly different values

Discussion

Brassica edible oil having high erucic acid causes myocardial infarction and increased blood cholesterol levels in human beings (Renard and Mcgregor 1992; Mortuza et al. 2006). The erucic acid content in *Brassica* seed oil is governed by *FAE1* gene encoding β-ketoacyl-CoA synthase (KCS) enzyme that catalyzes the synthesis of very long-chain monounsaturated fatty acids (VLCMFA), a major constituent of seed oil (Lühs et al. 1999). There are two independent genes *FAE1.1* and *FAE1.2* in allotetraploids *B. napus* and *B. juncea* (Gupta et al. 2004; Bhatiya and Alok 2014; Yan et al. 2015) controlling the erucic acid in additive gene interaction. The *FAE1* gene encodes for a 507 amino acids long polypeptide chain and shows high similarity between its alleles in *B. juncea* (97%), *B. napus* (96%), *B. rapa* (96%), *B. oleracea* (96%), and *Arabidopsis* (86%) (Zeng and Cheng 2014).

There are several SNPs present in the CDS of *FAE1* gene. Yan et al. (2015) identified 26 SNPs in *B. rapa* genotypes which led to 13 amino acid changes. In previous studies, it was reported that the SNPs in the CDS may lead to the change in the erucic acid content (Katavic et al. 2002; Hu et al. 2009; Wang et al. 2010; Wu et al. 2008). Saini et al. (2016) also reported the complete co-segregation between SNPs at positions 591 and 1265 in CDS of *FAE1.1* gene and at position 237 in the case of *FAE1.2* with erucic acid content. Using these SNPs, CAPS markers were developed and successfully used in *B. juncea (*Saini et al. 2016) for differentiating between LEA and HEA lines. In the present study also, no disequilibrium or recombination was found between the biochemical phenotype and genotype studied using those CAPS markers of *B. rapa*, *B. nigra*, *B. napus*, and *B. juncea* accessions. On the contrary, Zeng and Cheng (2014) observed that the low erucic acid is due to DNA methylation in the promoter region rather than the variation in CDS, in yellow mustard (*S. alba*). In the present study, the *FAE1* gene transcription profile was consistent with erucic acid content as the transcript level in mature embryo was upregulated (400 folds) in HEA genotype, while the similar upregulation in *FAE1* transcript was absent in LEA genotype. Yan et al. (2015) also observed the similar types of results in *B. rapa* but Hu et al. (2009) and Wu et al. (2008) got the contradictory result in *B. napus*. However, Li et al. (2017) found that swapping coding regions rather than promoters resulted in significant increase in the erucic acid (1.73 vs. 1.24) indicating that sequence variations in coding regions may be the predominant mechanism by which *FAE1* expression regulates the trait rather than promoter methylation.

The upstream region of about 1.3 kb of CDS is the putative promoter site of *FAE1* gene. The *FAE1* promoter is a seed-specific promoter (Zeng and Cheng 2014; Chiron et al. 2015) and also phylogenetically conserved in related species of *Brassica*. Among the diploid species, *B. rapa*, *B. nigra*, *B. oleracea*, and *S. alba*, the putative promoter region consists of approximately1-kb

long upstream sequence. About 600-bp region proximal to CDS is conserved among the above species with about 84% similarity, although the CDS reveals approximately 95% similarity. The *B. oleracea* and *Capsella rubella* promoter regions were found to be 48.7% similar compared to 84.9% homology between the coding regions (Li et al. 2017). The promoter sequence of *B. rapa* and *B. nigra* which are the sources of paralogs of *FAE1.1* and *FAE1.2* gene in *B. juncea* showed only 98% sequence homology, respectively.

The promoter sequence of *FAE1.1* gene showed about 95% similarity between LEA and HEA *B. juncea* genotypes. There are several SNPs and InDels between HEA and LEA throughout the promoter sequence. Association analysis revealed that the two InDels in AT-rich regions of *FAE1.1* promoter are significantly associated with erucic acid content. In *B. rapa*, AT-rich deletion in the promoter region was also associated with the low erucic acid content (Yan et al. 2015). The AT-rich sequences in promoter region act as enhancer elements in tissue-specific expression in higher plants (Sandhu et al. 1998) and also facilitate the formation of regulatory complexes that control transcription and recombination through protein–protein and protein–DNA interactions (Grasser 2003). Fukai et al. (2019) also reported the insertion of LTR region 362 bps in the upstream region of FAE CDS in *B rapa*.

In the case of *FAE1.2* gene, along with SNP and InDels, a transposable element Tc1/Mariner transposon is integrated into the promoter region in LEA genotypes. This insertion was also shown to be negatively associated with the erucic content in *B. juncea*. Insertions of DNA transposons or retrotransposons within or near structural genes modify or abolish transcription of genes (Hollister and Gaut 2009; Wang and Perry 2013). In *Sinapis alba*, the insertion of *PIF/Harbinger*-like DNA transposon in *FAE1* promoter region lowers the erucic acid in the oil (Zeng and Cheng 2014).

The allelic variation on the basis of the length polymorphism in the promoter region was exploited to develop the polymorphic markers used in the breeding program to develop the low erucic genotypes in *B. juncea*. These markers were validated in diverse genotypes and revealed the complete association between genotyping by CAPS (Saini et al. 2016) and biochemical phenotype. A primer set was able to capture the length polymorphism in the case of *FAE1.1* gene. However, in *FAE1.2*, the primer set could differentiate the length polymorphism in homozygous plants but failed to amplify LEA alleles in the heterozygous condition and also in vitro bulk of LEA & HEA DNA. This could possibly be due to the hetero-duplex formation during the polymerase chain reaction. The DNA strands amplified from different homologous templates during the PCR reaction could cross-hybridize under annealing conditions and form hetero-duplex or make D loop DNA molecules (Michu et al. 2010; Kalle et al. 2014). To overcome this difficulty, a single primer was designed against the transposable element and a set consisting of three primers was used in case of *FAE1.2* gene. The validation and association of these primers with contrasting phenotypes were carried out using BC_1F_2 population derived from Pusa Mustard 30 × Pusa Bold. The genotype of the backcrosses population perfectly co-segregated with the markers, sans any recombination indicating the reliability of these markers.

It is evident that the markers developed using the variations present in the promoter region in FAE1 loci are highly efficient. Use of these PCR-based markers in the marker-assisted selection for development of LEA *Brassica* genotypes will thus reduce time and cost, and will improve the accuracy.

Authors' contribution SV, NS, and DK conceived the project. All co-authors, NS, YP, NSH, MK, SY, RS, and SS were involved in executing the experiments. SV, NS, and YP were involved in the drafting of the manuscript. NSH, RB, and DK revised the final manuscript.

Funding The authors thank the Indian Council of Agricultural Research, Govt. of India, for the financial assistance to carry out this research.

References

Abbadi A, Leckband G (2011) Rapeseed breeding for oil content quality and sustainability. Eur J Lipid Sci Technol 113:1198–1206. https://doi.org/10.1002/ejlt.201100063

Bhatiya V, Alok A (2014) Molecular marker analysis of the linkage drag around the *FAE1* loci of *Brassica juncea* during conventional back cross breeding. J Crop Sci Biotechnol 17: 147–154. https://doi.org/10.1007/s12892-013-0135-1

Chiron H, Wilmer J, Lucas MO, Nesi N, Delseny M, Devic M, Roscoe TJ (2015) Regulation of FATTY ACID ELONGATION1 expression in embryonic and vascular tissues of *Brassica napus*. Plant Mol Biol 88:65–83. https://doi.org/10.1007/s11103-015-0309-y

Doyle JJ, Doyle JL (1990) Isolation of plant DNA from fresh tissue. Focus 12:13–15

Fukai E, Karim MM, Shea DJ, Tonu NN, Falk KC, Funaki T, Okazaki K (2019) An LTR retrotransposon insertion was the

cause of world's first low erucic acid *Brassica rapa* oilseed cultivar. Mol Breed 39(15). https://doi.org/10.1007/s11032-018-0916-9

Garcés-Claver A, Fellman SM, Gil-Ortega R, Jahn M, Arnedo-Andrés MS (2007) Identification validation and survey of a single nucleotide polymorphism (SNP) associated with pungency in *Capsicum* spp. Theor Appl Genet 115:907–916. https://doi.org/10.1007/s00122-007-0617-y

Grasser KD (2003) Chromatin-associated HMGA and HMGB proteins versatile co-regulators of DNA-dependent processes. Plant Mol Biol 53:281–295. https://doi.org/10.1023/B,PLAN.0000007002.99408.ba

Gupta V, Mukhopadhyay A, Arumugam N, Sodhi YS, Pental D, Pradhan AK (2004) Molecular tagging of erucic acid trait in oilseed mustard (*Brassica juncea*) by QTL mapping and single nucleotide polymorphisms in FAE1 gene. Theor Appl Genet 108:743–749. https://doi.org/10.1007/s00122-003-1481-z

Hollister JD, Gaut BS (2009) Epigenetic silencing of transposable elements a trade-off between reduced transposition and deleterious effects on neighboring gene expression. Genome Res 19:1419–1428. https://doi.org/10.1101/gr.091678.109

Hu Y, Wu G, Cao Y, Wu Y, Xiao L, Li X, Lu C (2009) Breeding response of transcript profiling in developing seeds of *Brassica napus*. BMC Mol Biol 10:49–65. https://doi.org/10.1186/1471-2199-10-49

Jagannath A, Sodhi YS, Gupta V, Mukhopadhyay A, Arumugam N, Singh I, Rohatgi S, Burma PK, Pradhan AK, Pental D (2011) Eliminating expression of erucic acid-encoding loci allows the identification of 'hidden' QTL contributing to oil quality fractions and oil content in *Brassica juncea* (Indian mustard). Theor Appl Genet 122:1091–1103. https://doi.org/10.1007/s00122-010-1515-2

James DW, Lim E, Keller J, Plooy I, Ralston E, Dooner HK (1995) Directed tagging of the *Arabidopsis fatty acid elongation 1* (*FAE1*) gene with the maize transposon activator. Plant Cell 7:309–319. https://doi.org/10.1105/tpc.7.3.309

Kalle E, Kubista M, Rensing C (2014) Multi-template polymerase chain reaction. Biomol Detect Quantif 2:11–29. https://doi.org/10.1016/j.bdq.2014.11.002

Katavic V, Mietkiewska E, Barton DL, Giblin EM, Reed DW, Taylor DC (2002) Restoring enzyme activity in nonfunctional low erucic acid Brassica napus fatty acid elongase 1 by a single amino acid substitution. FEBS J 269:5625–5631. https://doi.org/10.1046/j.1432-1033.2002.03270.x

Kirk JTO, Oram RN (1978) Mustard as possible oil and protein crops for Australia. J Aust Inst Agric Sci 44:143–156

Li D, Lei Z, Xue J, Zhou G, Hang Y, Sun X (2017) Regulation of FATTY ACID ELONGATION1 expression and production in *Brassica oleracea* and *Capsella rubella*. Planta 246:763–778. https://doi.org/10.1007/s00425-017-2731-7

Lühs WW, Voss A, Seyis F, Friedt W (1999) Molecular genetics of erucic acid content in the genus *Brassica*. In Proc 10th Intern Rapeseed Congr 26–29

Mag TK (1983) Canola oil processing in Canada. J Am Oil Chem Soc 60:380–384. https://doi.org/10.1007/BF02543522

Michu E, Mráčková M, Vyskot B, Žlůvová J (2010) Reduction of heteroduplex formation in PCR amplification. Biol Plant 54:173–176. https://doi.org/10.1007/s10535-010-0029-8

Millar AA, Kunst L (1997) Very-long-chain fatty acid biosynthesis is controlled through the expression and specificity of the condensing enzyme. Plant J 12:121–131. https://doi.org/10.1046/j.1365-313X.1997.12010121.x

Mortuza MG, Dutta PC, Das ML (2006) Erucic acid content in some rapeseed/mustard cultivars developed in Bangladesh. J Sci Food Agric 86:135–139. https://doi.org/10.1002/jsfa.2301

Renard S, Mcgregor S (1992) Antithrombogenic effect of erucic acid poor rapeseed oil in the rats. Rev Fr Crop Cros 23:393–396

Roscoe TJ, Lessire R, Puyaubert J, Renard M, Delseny M (2001) Mutations in the fatty acid elongation 1 gene are associated with a loss of *L-ketoacyl-CoA synthase* activity in low erucic acid rapeseed. FEBS Lett 492:107–111. https://doi.org/10.1016/S0014-5793(01)02243-8

Saini N, Singh N, Kumar A, Vihan N, Yadav S, Vasudev S, Yadava DK (2016) Development and validation of functional CAPS markers for the FAE genes in *Brassica juncea* and their use in marker-assisted selection. Breed Sci 66:831–837. https://doi.org/10.1270/jsbbs.16132

Sandhu JS, Webster CI, Gray JC (1998) A/T-rich sequences act as quantitative enhancers of gene expression in transgenic tobacco and potato plants. Plant Mol Biol 37:885–896. https://doi.org/10.1023/A,1006051832213

Sujata V, Yadava DK, Divya M, Tanwar RS, Prabhu KV (2008) A simplified method for preparation of fatty acid methyl esters of *Brassica* oil. Indian J Genet 68:456–458

Wang F, Perry SE (2013) Identification of direct targets of FUSCA3 a key regulator of Arabidopsis seed development. Plant Physiol 161:1251–1264. https://doi.org/10.1104/pp.112.212282

Wang N, Shi L, Tian F, Ning H, Wu X, Long Y, Meng J (2010) Assessment of *FAE1* polymorphism in three *Brassica* species using EcoTILLING and their association with differences in seed erucic acid contents. BMC Plant Biol 10:137. https://doi.org/10.1186/1471-2229-10-137

Wu G, Wu Y, Xiao L, Li X, Lu C (2008) Zero erucic acid trait of rapeseed (*Brassica napus* L) results from a deletion of four base pairs in the fatty acid elongase 1 gene. Theor Appl Genet 116:491–499. https://doi.org/10.1007/s00122-007-0685-z

Yan G, Li D, Cai M, Gao G, Chen B, Xu K, Li J, Li F, Wang N, Qiao J, Li H, Zhang T, Wu X (2015) Characterisation of *FAE1* in the zero erucic acid germplasm of *Brassica rapa* L. Breed Sci 65:257–264. https://doi.org/10.1270/jsbbs.65.257

Zeng F, Cheng B (2014) Transposable element insertion and epigenetic modification cause the multiallelic variation in the expression of *FAE1* in *Sinapis alba*. Plant Cell 26:2648–2659. https://doi.org/10.1105/tpc.114.126631