

# SCIENTIFIC REPORTS

OPEN

## Mutations in the promoter, intron and CDS of two *FAD2* generate multiple alleles modulating linoleic acid level in yellow mustard

Fangqin Zeng, Vicky Roslinsky &amp; Bifang Cheng

Linoleic acid (C18:2) is an important polyunsaturated fatty acid in the seed oil of many crops. Here, we report that mutations in the promoter, intron and CDS of the *FAD2* genes *SalFAD2.LIA1* and *SalFAD2.LIA2* generate three alleles *LIA<sup>1a</sup>*, *LIA<sup>1b</sup>* and *lia<sup>1</sup>* and two alleles *LIA<sup>2</sup>* and *lia<sup>2</sup>*, respectively, controlling the C18:2 variation (4.4–32.7%) in yellow mustard. The allelic effect on increasing C18:2 content is *LIA<sup>1a</sup>* > *LIA<sup>1b</sup>* > *lia<sup>1</sup>*, *LIA<sup>2</sup>* > *lia<sup>2</sup>*, and *LIA<sup>1a</sup>* > *LIA<sup>2</sup>*. The five *FAD2* alleles each contain two exons, one intron and a promoter adjacent to exon 1. *LIA<sup>1a</sup>* has a 1152 bp CDS, a 1221 bp intron with promoter function and a 607 bp promoter. Compared with *LIA<sup>1a</sup>*, the intron of *LIA<sup>1b</sup>* has reduced promoter activity and that of *LIA<sup>2</sup>* and *lia<sup>2</sup>* has no promoter function due to extensive SNP and indel mutations. *lia<sup>1</sup>* differed from *LIA<sup>1b</sup>* by having an insertion of 1223 bp retrotransposon in its intron. *lia<sup>2</sup>* with mutations in the promoter has reduced promoter activity compared with *LIA<sup>2</sup>*. This study revealed that complex quantitative variation of trait phenotype in plants could be modulated by multiple alleles of oligogenic loci resulting from mutations in the regulatory region and CDS.

Polyunsaturated fatty acids including linoleic (C18:2) and linolenic (C18:3) acids are the major components of seed storage lipids in higher plants and important factors in determining the qualities of edible oil such as oxidative stability<sup>1</sup> and nutritional value<sup>2</sup>. Biosynthesis of linoleic acid is catalyzed by the microsomal delta-12 fatty acid desaturase (*FAD2*) localized in the endoplasmic reticulum<sup>3</sup>. The *FAD2* gene in *Arabidopsis thaliana* was isolated from mutants created by T-DNA insertions<sup>4</sup>. Subsequent investigations have identified one or more *FAD2* copies in various crops such as cabbage (*Brassica oleracea*) and turnip (*B. rapa*)<sup>5</sup>, rapeseed (*B. napus*)<sup>6</sup>, cotton (*Gossypium hirsutum*)<sup>7</sup>, sesame (*Sesamum indicum*)<sup>8</sup>, olive (*Olea europaea*)<sup>9</sup> and grape (*Vitis labrusca*)<sup>10</sup>. The *FAD2* gene contains two exons, one intron embedded within the 5' UTR and a promoter<sup>5–7, 10, 11</sup>. The intron of *FAD2* has promoter function and can regulate the expression level of *FAD2*<sup>6, 11</sup>. The protein encoded by *FAD2* contains six transmembrane domains and three histidine boxes (H box) harboring eight iron-binding histidines in *A. thaliana*<sup>4</sup>. The histidines appear to be crucial for proper enzymatic function, since substitution of a histidine with a different amino acid disrupts desaturase function<sup>12, 13</sup>.

It has been reported that mutations in the coding DNA sequence (CDS) of *FAD2* are correlated with the increase of oleic acid (C18:1) content in *B. napus*<sup>14–16</sup> and *B. rapa*<sup>17</sup>. The high oleic acid (~77%) variant DMS100, developed through ethyl methane sulfonate (EMS) mutagenesis, was due to a single nucleotide mutation that resulted in the occurrence of a stop codon (TAG) leading to premature termination of the open reading frame of *FAD2* in *B. napus*<sup>14</sup>. However, only partial coding DNA sequences (nucleotide 148–1128) of the *FAD2* were cloned from DMS 100 and the wild-type line Quantum in this study. Yang *et al.* (2012) successfully cloned four copies of *FAD2* genes from the high C18:1 (~78%) variant SW Hiskory and the wild-type line JA177 in *B. napus*. Sequence alignment indicated that a 4 bp insertion at the position 567–568th base pair of the *FAD2* gene *BnaA.FAD2.a* was responsible for the high C18:1 content in SW Hiskory<sup>15</sup>. Interestingly, *BnaA.FAD2.a* of the high C18:1 (~75%) variant Cabriolet was also non-functional, but resulted from a 1 bp deletion, leading to a frame shift and a truncated protein in *B. napus*<sup>16</sup>. The high C18:1 content of line Jo4072 could be resulted from the transition at nucleotide 484 in the CDS of *FAD2* in *B. rapa*<sup>17</sup>. It is worth-noting that the function of the encoded enzyme, the

Agriculture and Agri-Food Canada, Saskatoon Research and Development Centre, 107 Science Place, Saskatoon, SK, S7N 0X2, Canada. Correspondence and requests for materials should be addressed to B.C. (email: [bifang.cheng@agr.gc.ca](mailto:bifang.cheng@agr.gc.ca))

Parental lines	Generation	C18:1 (% of total fatty acids)	C18:2 (% of total fatty acids)	C18:3 (% of total fatty acids)	C22:1 (% of total fatty acids)
Y1798	S <sub>4</sub>	27.3 ± 2.3	4.2 ± 0.1	13.0 ± 0.4	36.3 ± 2.4
Y514	DH	67.6 ± 0.5	12.5 ± 0.2	10.8 ± 0.5	0.2 ± 0.1
Y1801	S <sub>5</sub>	21.7 ± 0.7	31.5 ± 0.9	12.5 ± 0.5	16.1 ± 0.3
Y1798 (low) × Y1801 (high)	F <sub>1</sub>	29.3 ± 0.6	10.2 ± 0.2** <sup>1</sup>	14.7 ± 0.4	26.2 ± 0.3
Mid-parent value of Y1798 and Y801		24.5	18.6	12.8	26.2
Y1798 (low) × Y514 (medium)	F <sub>1</sub>	32.0 ± 0.4	6.9 ± 0.2** <sup>2</sup>	13.3 ± 0.2	26.6 ± 0.3
Mid-parent value of Y1798 and Y514		47.8	8.8	11.9	18.3
Y1801 (high) × Y514 (medium)	F <sub>1</sub>	30.2 ± 0.7	21.4 ± 0.4	17.8 ± 0.3	11.1 ± 0.2
Mid-parent value of Y1801 and Y514		44.7	23.0	11.7	8.2

**Table 1.** Fatty acid profile of the parental lines Y1798, Y514 and Y1801, F<sub>1</sub> seeds, and mid-parent value in yellow mustard. Fatty acid content is expressed as mean value ± standard deviation. \*\*Statistical significant difference at  $p = 0.01$  level. <sup>1</sup>Comparison between F<sub>1</sub> and mid-parent value of Y1798 and Y1801; <sup>2</sup>Comparison of F<sub>1</sub> and mid-parent value of Y1798 and Y514.

transcriptional level and the promoter activity of different *FAD2* alleles were not characterized and the introns of *FAD2* were not found in these studies.

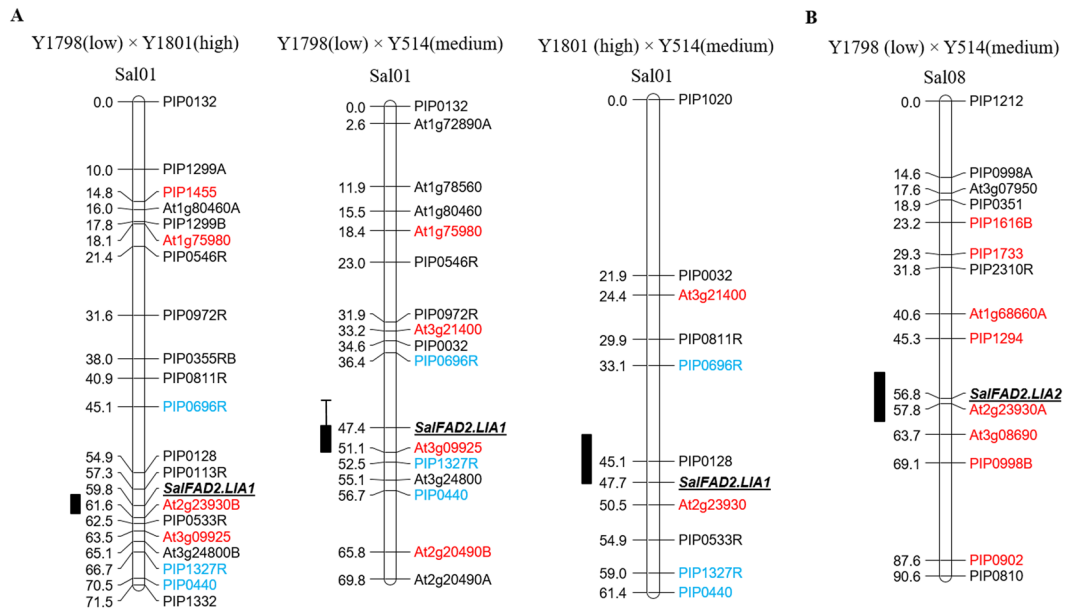
Condiment yellow mustard (*Sinapis alba* L.) is phylogenetically related to *Brassica* species<sup>18</sup> and is an obligate outcrossing crop. Open-pollinated (OP) population varieties of this crop comprise great genetic variation. Various plant morphological types, fatty acid and glucosinolate variants have been isolated via inbreeding of the OP varieties in yellow mustard<sup>19–21</sup>. The molecular mechanism underlying the natural occurring erucic variants in yellow mustard is revealed to be different from that reported in *B. napus*. The erucic acid variants have resulted from SNP/indel mutations in the CDS of fatty acid elongation 1 (*FAE1*) in *B. napus*<sup>22–26</sup>. However, transposable element insertions and epigenetic modification in the *FAE1* had led to the occurrence of multi-alleles in yellow mustard<sup>27</sup>. Linoleic acid (C18:2) content ranged from 6.2% to 14.2% in yellow mustard germplasm<sup>28</sup>. Recently, three lines Y1798, Y514 and Y1801 with low (average: 4.2%, range: 3.5–5.3%), medium (average: 12.5%, range: 12.0–13.0%) and high (average: 31.5%, range: 28.7–37.7%) C18:2 contents, respectively, have been successfully developed in this species<sup>29,30</sup>. In the present study, the CDS, promoter and intron of *FAD2* from each of the three lines Y1801, Y514 and Y1798 have been molecularly and functionally characterized. Here, we report that mutations in the promoter, intron and CDS of two *FAD2* generate multiple alleles modulating the quantitative variation of C18:2 content in yellow mustard.

## Results

**Mapping QTLs for C18:2 content.** The average C18:2 contents of the F<sub>1</sub> seeds derived from the crosses of Y1798 (low) × Y1801 (high) and Y1798 (low) × Y514 (medium) were 10.2% and 6.9%, which were significantly lower than the mid-parent values of 18.6% ( $t = 9.80, p < 0.01$ ) and 8.8% ( $t = 2.60, p < 0.01$ ), respectively (Table 1; Supplementary Fig. S1), suggesting a partial dominance of the low over both high and medium C18:2 contents. The F<sub>1</sub> seeds of Y1801 (high) × Y514 (medium) had an average C18:2 content of 21.4% which was similar with the mid-parent value of 23.0% ( $t = 0.93, p = 0.37$ ) (Table 1; Supplementary Fig. S1). The F<sub>2</sub> seeds of each of the three crosses showed a continuous frequency distribution in the C18:2 content (Supplementary Fig. S2) and were not possible to be classified into discrete groups. Therefore, QTL mapping was used to identify the QTLs controlling C18:2 content.

One hundred thirty, one hundred ten and one hundred fourteen polymorphic ILP markers and the allele-specific markers for the *FAE1* gene and the *FAD3* genes *SalFAD3.LA1* and *SalFAD3.LA2* were used to genotype individual plants of the F<sub>2</sub> populations of Y1798 (low) × Y1801 (high), Y1798 (low) × Y514 (medium) and Y1801 (high) × Y514 (medium). Twelve linkage groups were constructed using the polymorphic markers genotyped in each of the three crosses. Based on the common ILP markers, the 12 linkage groups corresponded to Sal01 to Sal12 of the constructed *S. alba* map by Javidfar and Cheng (2013)<sup>31</sup>. The *FAE1* gene was mapped to Sal03 and the *FAD3* genes *SalFAD3.LA1* to Sal02 and *SalFAD3.LA2* to Sal10 (Supplementary Fig. S3A–C).

Three QTLs for C18:2 content, two QTLs for C18:3 content and one QTL for C22:1 content were identified in the F<sub>2</sub> population of Y1798 (low) × Y1801 (high) (Fig. 1, Supplementary Table 2 and Supplementary Fig. S3A). The three C18:2 QTLs explained 43.2%, 42.8% and 5.6% of the phenotypic variation and mapped to the linkage groups Sal01, Sal02 and Sal03, respectively. The C18:2 QTL (LOD = 40.63) on Sal01 was located between the markers PIP0113R and At2g23930B (Fig. 1A). The C18:2 QTL (LOD = 40.44) on Sal02 co-localized with the QTL for C18:3 content and the *FAD3* gene *SalFAD3.LA1* (Supplementary Fig. S3A). The C18:2 QTL (LOD = 9.99) on Sal03 shared the same region with the C22:1 QTL and the *FAE1* gene (Supplementary Fig. S3A). In the F<sub>2</sub> population of Y1798 (low) × Y514 (medium), three C18:2 QTLs, three C18:3 QTLs and one C22:1 QTL were detected.

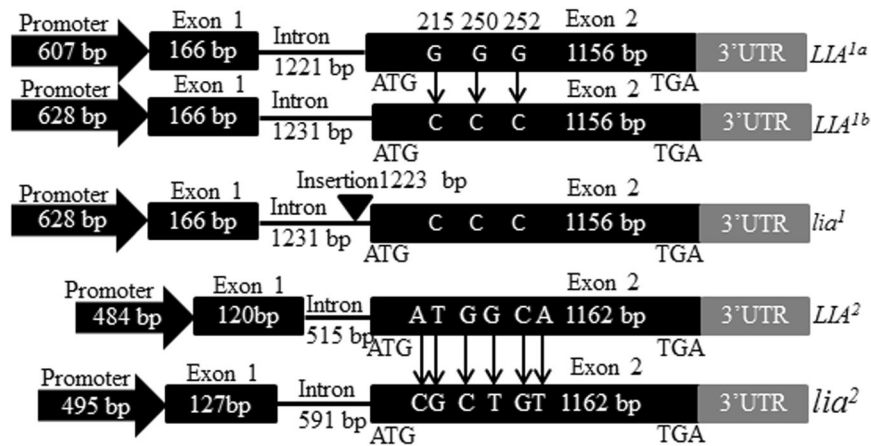


**Figure 1.** Mapping QTLs controlling C18:2 content in yellow mustard. **(A)** The C18:2 QTL in Sal01 is located between PIP0113R and At2g23930B in Y1798 (low) × Y1801 (high), between PIP0696R and At3g09925 in Y1798 (low) × Y514 (medium), and between PIP0696R and At2g23930 in Y1801 (high) × Y514 (medium). **(B)** The C18:2 QTL in Sal08 is located between PIP1294 and At3g08690 in Y1798 (low) × Y514 (medium). 1-LOD and 2-LOD supporting intervals of each C18:2 QTL were marked by thick and thin bars, respectively. Markers in blue font were detected in all three crosses. Markers in red font were mapped to the same linkage groups as reported by Javidfar and Cheng (2013)<sup>31</sup>. The *SalFAD2.LIA1* and *SalFAD2.LIA2* genes co-localized with their C18:2 QTL peaks in the linkage groups Sal01 and Sal08, respectively.

The three C18:2 QTLs, responsible for 47.3%, 24.4% and 9.5% of the phenotypic variation, were assigned to Sal01, Sal03 and Sal08, respectively (Fig. 1, Supplementary Table 2 and Supplementary Fig. S3B). The C18:2 QTL (LOD = 25.44) on Sal01 was located between the markers PIP0696R and At3g09925 (Fig. 1A). The C18:2 QTL on Sal03 co-localized with the C22:1 QTL and the *FAE1* gene (Supplementary Fig. S3B). The C18:2 QTL (LOD = 3.47) on Sal08 was located between the markers PIP1294 and At3g08690 (Fig. 1B). Two C18:2 QTLs, two C18:3 QTLs and one C22:1 QTL were revealed in the F<sub>2</sub> population of Y1801 (high) × Y514 (medium) (Fig. 1, Supplementary Table 2 and Supplementary Fig. S3C). The C18:2 QTL (LOD = 14.40), accounting for 42.6% of the C18:2 content variation, was mapped to the same region as the C18:3 QTL and *FAD3* gene *SalFAD3.LA1* on Sal02 (Supplementary Fig. S3C). To eliminate the confounding effect of C18:3 variation, 22 F<sub>2</sub> plants carrying the dominant homozygous *SalFAD3.LA1* alleles *LIA<sup>1</sup>LIA<sup>1</sup>* were removed from the F<sub>2</sub> population. QTL analysis using the remaining 102 F<sub>2</sub> plants identified one C18:2 QTL (LOD = 3.54) on Sal01, responsible for 14.9% of the C18:2 content variation (Fig. 1A) in addition to the C18:3 QTL on Sal02. This indicated that the effect of the C18:3 QTL on C18:2 content is larger than the C18:2 QTL.

The C18:2 QTL on Sal02 co-localized with the *FAD3* gene *SalFAD3.LA1* and that on Sal03 with the *FAE1* gene. Therefore, only the C18:2 QTLs on Sal01 and Sal08 implied the presence of *FAD2* genes. The *FAD2* gene on Sal01 was referred to as *SalFAD2.LIA1*, while that on Sal08 as *SalFAD2.LIA2*. The C18:2 QTL on Sal01 was detected in the three crosses Y1798 (low) × Y1801 (high), Y1798 (low) × Y514 (medium) and Y1801 (high) × Y514 (medium), indicating that Y1801, Y514 and Y1798 carried different alleles, designated as *LIA<sup>1a</sup>*, *LIA<sup>1b</sup>* and *lia<sup>1</sup>*, respectively, at the *SalFAD2.LIA1* locus. The C18:2 QTL on Sal08 was identified in the cross of Y1798 (low) × Y514 (medium), but not in Y1801 (high) × Y514 (medium), suggesting that Y1801 and Y514 harboured the same allele, designated as *LIA<sup>2</sup>*, while Y1798 had the recessive allele *lia<sup>2</sup>* at the *SalFAD2.LIA2* locus. Thus, the C18:2 genotypes of Y1801, Y514 and Y1798 were *LIA<sup>1a</sup>LIA<sup>1a</sup>LIA<sup>2</sup>LIA<sup>2</sup>*, *LIA<sup>1b</sup>LIA<sup>1b</sup>LIA<sup>2</sup>LIA<sup>2</sup>* and *lia<sup>1</sup>lia<sup>1</sup>lia<sup>2</sup>lia<sup>2</sup>*, respectively.

**Cloning of the *SalFAD2.LIA1* alleles *LIA<sup>1a</sup>*, *LIA<sup>1b</sup>* and *lia<sup>1</sup>* and the *SalFAD2.LIA2* alleles *LIA<sup>2</sup>* and *lia<sup>2</sup>*.** The coding DNA sequence (CDS) of *LIA<sup>1a</sup>* and *LIA<sup>2</sup>* were successfully cloned from Y1801 (*LIA<sup>1a</sup>LIA<sup>1a</sup>LIA<sup>2</sup>LIA<sup>2</sup>*), while the CDS of *LIA<sup>1b</sup>* and *LIA<sup>2</sup>* from Y514 (*LIA<sup>1b</sup>LIA<sup>1b</sup>LIA<sup>2</sup>LIA<sup>2</sup>*) and the CDS of *lia<sup>1</sup>* and *lia<sup>2</sup>* from Y1798 (*lia<sup>1</sup>lia<sup>1</sup>lia<sup>2</sup>lia<sup>2</sup>*) using primer pair No 1 (Supplementary Table S1; Supplementary Fig. S4). *LIA<sup>1a</sup>*, *LIA<sup>1b</sup>* and *lia<sup>1</sup>* had a CDS of 1152 bp encoding a polypeptide of 383 amino acids. The CDS of *LIA<sup>1b</sup>* and *lia<sup>1</sup>* shared the same nucleotide sequence. Sequence alignment with *LIA<sup>1a</sup>* allowed the identification of fifteen point mutations at positions 99, 156, 171, 215, 250, 252, 396, 534, 573, 615, 681, 684, 735, 957 and 1032 in the CDS of *LIA<sup>1b</sup>* and *lia<sup>1</sup>*. The mutations at positions 215, 250 and 252 (G of *LIA<sup>1a</sup>* to C of *LIA<sup>1b</sup>* and *lia<sup>1</sup>*) led to the two amino acid changes: the threonine and phenylalanine residues at position 72 and 84 in the protein encoded by *LIA<sup>1a</sup>* were substituted by the serine and valine residues in the protein encoded by *LIA<sup>1b</sup>* and *lia<sup>1</sup>* (Fig. 2; Supplementary Fig. S5). The cloned CDS of *LIA<sup>2</sup>* from Y1801 and Y514 were identical in size and nucleotide sequence. *LIA<sup>2</sup>* and



**Figure 2.** Schematic representation of the structure of the *SalFAD2.LIA1* alleles *LIA1<sup>a</sup>*, *LIA1<sup>b</sup>*, *lia1* and the *SalFAD2.LIA2* alleles *LIA2* and *lia2*. Black rectangle: Exon; Straight line: Intron; Solid black arrow: Promoter; Grey rectangle: 768 bp 3' UTR; Vertical long arrow: Point mutation; Inverted triangle: Insertion.

*lia2* had a CDS of 1158 bp encoding a polypeptide of 385 amino acids. Twelve point mutations at positions 34, 78, 125, 174, 187, 363, 447, 450, 506, 597, 616 and 642 were observed in the CDS of *lia2* when compared with that of *LIA2*. The mutations at positions 34, 125, 174, 187, 506, and 642 caused amino acid changes; the threonine, proline, cysteine, serine, glycine and asparagine residues at positions 12, 42, 58, 63, 169 and 206 in the protein sequence encoded by *LIA2* were substituted by serine, histidine, tryptophan, alanine, alanine and aspartic acid residues in the protein encoded by *lia2* (Supplementary Fig. S5).

Sequence alignment of the CDS of *LIA1<sup>a</sup>* with that of *LIA2* indicated that *LIA1<sup>a</sup>* harboured a 3 bp deletion at position 516 and a 3 bp deletion at position 613, which resulted in the loss of the asparagine residue at position 175 and the tyrosine residue at position 204 in the polypeptide (Supplementary Fig. S5). In addition, 104 SNPs were observed in the CDS of *LIA1<sup>a</sup>* and *LIA2*, of which 27 SNPs led to amino acid changes in the first, second, third, fifth and sixth transmembrane domains (Supplementary Fig. S5).

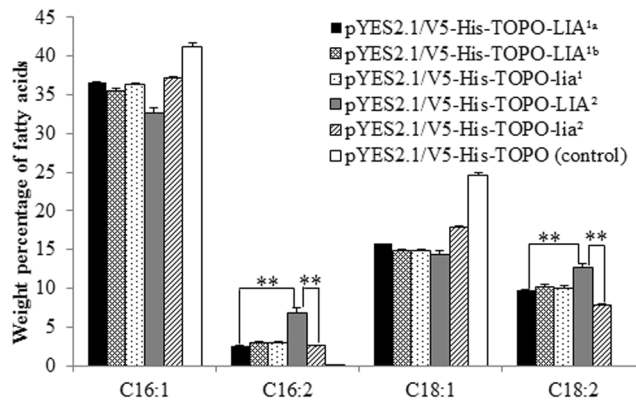
DNA fragments comprising the 736 bp downstream sequences from the translation stop codon were cloned for *LIA1<sup>a</sup>*, *LIA1<sup>b</sup>*, *lia1*, *LIA2* and *lia2*, respectively, using primer pair No 2 (Supplementary Table S1; Supplementary Fig. S4). These DNA fragments had the same nucleotide sequence.

The 5' flanking sequences of *LIA1<sup>a</sup>*, *LIA1<sup>b</sup>* and *lia1* were cloned by two rounds of PCR walking using primer pairs No 3 and No 4 (Supplementary Table S1; Supplementary Fig. S4). The cloned 5' flanking fragments of *LIA1<sup>a</sup>*, *LIA1<sup>b</sup>* and *lia1* were 671 bp, 681 bp and 1909 bp, respectively. BLAST analysis of the cloned upstream sequences of *LIA1<sup>a</sup>*, *LIA1<sup>b</sup>* and *lia1* against the NCBI nucleotide sequence database revealed that they had high similarity to the intron of *BnFAD2* (KF038144) in *B. napus* with the E value being  $7e^{-105}$ ,  $3e^{-77}$  and  $3e^{-81}$ , respectively. Therefore, the promoter sequence of *BnFAD2* (KF038144) was used to design primer pair No 5 (Supplementary Table S1; Supplementary Fig. S4) to clone the promoters for *LIA1<sup>a</sup>*, *LIA1<sup>b</sup>* and *lia1*. The cloned promoter of *LIA1<sup>a</sup>* was 607 bp, while those of *LIA1<sup>b</sup>* and *lia1* were 628 bp. The gap between the promoter region and the 5' flanking fragment was filled with a 720 bp fragment for *LIA1<sup>a</sup>*, a 724 bp fragment for *LIA1<sup>b</sup>* and a 719 bp fragment for *lia1*, obtained using primer pair No 6 (Supplementary Table S1; Supplementary Fig. S4). Sequence alignments of the putative promoter and the 5' flanking fragments obtained using primer pairs No 3, 4, 5 and 6 revealed that the cloned 5' upstream sequences of *LIA1<sup>a</sup>*, *LIA1<sup>b</sup>* and *lia1* are 1998 bp, 2019 bp and 3252 bp, respectively.

The 5' flanking sequences of *LIA2* and *lia2* were cloned by two rounds of PCR walking using primer pairs No 7 and No 8 (Supplementary Table S1; Supplementary Fig. S4). The cloned 5' flanking fragments of *LIA2* from Y1801 and Y514 were 1123 bp and identical in sequence, while that of *lia2* from Y1798 was 1585 bp. BLAST analysis of the 1123 bp upstream sequence of *LIA2* and the 1585 bp upstream sequence of *lia2* against the NCBI nucleotide sequence database revealed that they had high similarity with the promoter, first exon and intron sequences of *BjFAD2-1* (HM147243) in *B. juncea* with the E value of  $3e^{-15}$  and  $9e^{-11}$ , respectively.

Thus, the entire nucleotide sequence (5' upstream and coding regions) for each of the *FAD2* alleles *LIA1<sup>a</sup>*, *LIA1<sup>b</sup>*, *lia1*, *LIA2* and *lia2* with sizes of 3150 bp, 3181 bp, 4404 bp, 2281 bp and 2744 bp, respectively, was obtained by PCR walking. To validate these results, *LIA1<sup>a</sup>*, *LIA1<sup>b</sup>* and *lia1* were each cloned in its entirety using primer pair No 9, while *LIA2* and *lia2* using primer pair No 10 (Supplementary Table S1). As expected, the resulting DNA fragments of the *SalFAD2.LIA1* and *SalFAD2.LIA2* alleles had the same sizes (Supplementary Fig. S6) and nucleotide sequences as those obtained from PCR walking.

**The *SalFAD2.LIA1* and *SalFAD2.LIA2* alleles differ in their promoter, exons 1 and 2, and intron.** Sequence alignment with *BnFAD2* (KF038144) from *B. napus* indicated that the cloned 5' upstream sequences of *LIA1<sup>a</sup>*, *LIA1<sup>b</sup>* and *lia1* comprised the first exon. A 1322 bp cDNA fragment was thus successfully cloned for *LIA1<sup>a</sup>*, *LIA1<sup>b</sup>* and *lia1* using primer pair No 11 (Supplementary Table S1) designed based on the sequences of their deduced first exon and cloned CDS. Sequence alignment of the cDNA and genomic DNA sequences indicated that *LIA1<sup>a</sup>*, *LIA1<sup>b</sup>* and *lia1* each comprised a promoter, exon 1, an intron and exon 2, but differed in size and nucleotide sequence (Fig. 2). The promoter, exon 1, intron and exon 2 of *LIA1<sup>a</sup>* were 607 bp, 166 bp, 1221 bp and

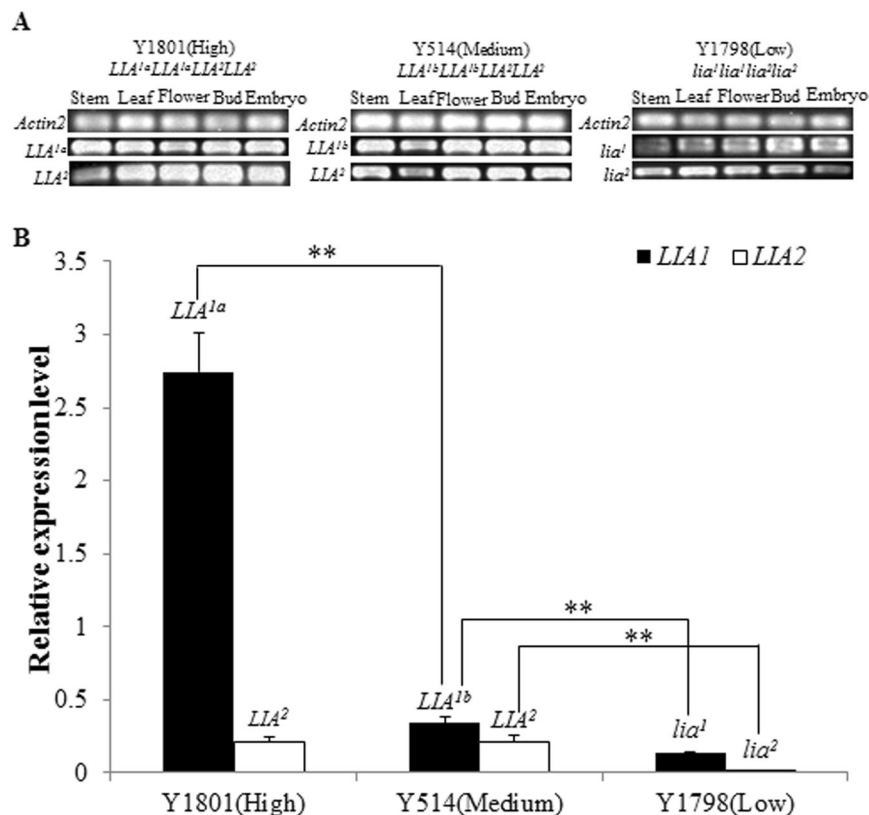


**Figure 3.** Heterologous expression of the *SalFAD2.LIA1* alleles *LIA<sup>1a</sup>*, *LIA<sup>1b</sup>* and *lia<sup>1</sup>*, and *SalFAD2.LIA2* alleles *LIA<sup>2</sup>* and *lia<sup>2</sup>* in Yeast. Gas chromatography analysis of fatty acid composition of yeast cells containing the construct pYES2.1/V5-His-TOPO-*LIA<sup>1a</sup>* with *LIA<sup>1a</sup>* coding DNA sequence (CDS), pYES2.1/V5-His-TOPO-*LIA<sup>1b</sup>* with *LIA<sup>1b</sup>* CDS, pYES2.1/V5-His-TOPO-*lia<sup>1</sup>* with *lia<sup>1</sup>* CDS, pYES2.1/V5-His-TOPO-*LIA<sup>2</sup>* with *LIA<sup>2</sup>* CDS, pYES2.1/V5-His-TOPO-*lia<sup>2</sup>* with *lia<sup>2</sup>* CDS and the empty vector pYES2.1/V5-His-TOPO. \*\*Statistical significant difference at  $p = 0.01$  level.

1156 bp, respectively. *LIA<sup>1b</sup>* and *lia<sup>1</sup>* shared the same promoter with 628 bp and the first exon of 166 bp and the second exon of 1156 bp. The intron of *lia<sup>1</sup>* was different from that (1231 bp) of *LIA<sup>1b</sup>* by having a 1223 bp insertion at position  $-26$  bp in the 5' upstream region from the translation start codon (Fig. 2; Supplementary Fig. S7). Sequence alignment with the promoter of *LIA<sup>1a</sup>* revealed that the promoter sequences of *LIA<sup>1b</sup>* and *lia<sup>1</sup>* harboured insertions of 12 bp, 4 bp, 4 bp, 1 bp, 1 bp and 3 bp at position 25, 171, 392, 494, 522 and 548, and deletions of 3 bp and 1 bp at positions 66 and 369 as well as 9 point mutations at positions 57, 134, 135, 337, 353, 404, 491, 534 and 546 (Supplementary Fig. S8A). Exon 1 of *LIA<sup>1b</sup>* and *lia<sup>1</sup>* harboured a 2 bp deletion at position 34 and a 2 bp insertion at position 153 as well as three point mutations at positions 25, 82 and 106 in comparison with that of *LIA<sup>1a</sup>* (Supplementary Fig. S8B). The exon 2 of *LIA<sup>1a</sup>*, *LIA<sup>1b</sup>* and *lia<sup>1</sup>* contained the CDS of 1152 bp. Sequence alignment with the intron of *LIA<sup>1a</sup>* revealed that the intron of *LIA<sup>1b</sup>* harboured insertions of 5 bp, 1 bp, 1 bp, 1 bp, 5 bp, 5 bp and 3 bp at positions 50, 71, 771, 830, 877, 1142 and 1192, and deletions of 1 bp at positions 135, 200, 271, 373, 421, 956, 1046 and 1087, a 3 bp deletion at position 445 as well as 49 point mutations (Supplementary Fig. S9). The 1223 bp insertion in the intron of *lia<sup>1</sup>* contained a coding region of 1044 bp that has a primer binding site (PBS) (Supplementary Fig. S10A) and encode a protein with 348 amino acids (Supplementary Fig. S10B). The protein contained the conserved domain of gag-polypeptide of LTR copia-type retrotransposon and exhibited 65% identity with putative reverse transcriptase (AAD11595) in *Arabidopsis* ( $E = 4e^{-103}$ ). Therefore, the 1223 bp insertion in the intron of *lia<sup>1</sup>* appears to be a retrotransposon.

Sequences alignment with *BjFAD2-1* (HM147243) from *B. juncea* indicated that the cloned 5' upstream sequences of the *SalFAD2.LIA2* alleles *LIA<sup>2</sup>* and *lia<sup>2</sup>* contained the first exon. The cDNA sequences of 1282 bp and 1289 bp for *LIA<sup>2</sup>* and *lia<sup>2</sup>*, respectively, were cloned using primer pair No 12 (Supplementary Table S1) designed based on the deduced first exon and the cloned CDSs. Alignment of the cDNA and genomic DNA sequences indicated that *LIA<sup>2</sup>* contained a promoter of 484 bp, the first exon of 120 bp, the second exon of 1162 bp and an intron of 515 bp, while *lia<sup>2</sup>* had a promoter of 495 bp, the first exon of 127 bp, the second exon of 1162 bp and an intron of 591 bp (Fig. 2). Sequences alignment with *LIA<sup>2</sup>* revealed that *lia<sup>2</sup>* harboured deletions of a 2 bp, 4 bp, 1 bp, 2 bp and 1 bp at positions 2, 184, 222, 274 and 420, and insertions of 1 bp, 1 bp, 7 bp, 10 bp and 1 bp at positions 312, 339, 353, 368 and 434 in the promoter (Supplementary Fig. S11A), and a 5 bp and a 2 bp insertion at positions 32 and 47 in the first exon (Supplementary Fig. S11B). The exon 2 of *LIA<sup>2</sup>* and *lia<sup>2</sup>* comprised the CDS of 1158 bp that differed in sequence (Fig. 2). The intron of *lia<sup>2</sup>* harboured 9 bp, 2 bp, 1 bp, 2 bp, 1 bp and 14 bp deletions at positions 76, 141, 154, 203, 532 and 546, and 17 bp, 9 bp, 17 bp, 15 bp, 9 bp, 29 bp, 8 bp and 1 bp insertions at positions 247, 276, 297, 381, 399, 413, 474 and 594, respectively, compared with the intron of *LIA<sup>2</sup>* (Supplementary Fig. S11C). Compared with *LIA<sup>1a</sup>*, *LIA<sup>2</sup>* harboured 103 SNPs, 5 insertions and 24 deletions in the promoter and 106 SNPs, one 5 bp insertion and 43 deletions in the intron (Supplementary Fig. S12).

**Heterologous expression of the *SalFAD2.LIA1* and *SalFAD2.LIA2* alleles in yeast.** Transgenic yeast cultures containing the empty construct pYES2.1/V5-His-TOPO produced the typical fatty acids, *i.e.* C16:0, C16:1, C18:0 and C18:1, found in untransformed cells (Fig. 3). Transgenic yeast cells harbouring the yeast expression vector pYES2.1/V5-His-TOPO-*LIA<sup>1a</sup>* carrying *LIA<sup>1a</sup>* coding sequence and pYES2.1/V5-His-TOPO-*LIA<sup>1b</sup>* (*lia<sup>1</sup>*) with *LIA<sup>1b</sup>* (*lia<sup>1</sup>*) coding sequence produced C16:2 and C18:2 fatty acids, and both C16:2 and C18:2 contents were not significantly different between cultures expressing *LIA<sup>1a</sup>* versus *LIA<sup>1b</sup>* (*lia<sup>1</sup>*) (Fig. 3). This result indicated that the cloned alleles *LIA<sup>1a</sup>* and *LIA<sup>1b</sup>* (*lia<sup>1</sup>*) encoded functional enzymes capable of desaturation of C16:1 to C16:2 and C18:1 to C18:2. Transgenic yeast cells harbouring the constructs pYES2.1/V5-His-TOPO-*LIA<sup>2</sup>* with *LIA<sup>2</sup>* coding sequence and pYES2.1/V5-His-TOPO-*lia<sup>2</sup>* with *lia<sup>2</sup>* coding sequence produced different amounts of C16:2 and C18:2 fatty acids. The C16:2 ( $6.5\% \pm 0.6\%$  SD) and C18:2 ( $12.6\% \pm 0.8\%$  SD) contents of yeast cells expressing *LIA<sup>2</sup>* were significantly higher than the C16:2 content ( $2.6\% \pm 0.2\%$  SD) ( $t = 6.49$ ,  $p < 0.01$ ) and C18:2 content ( $7.8\% \pm 0.4\%$  SD) ( $t = 8.79$ ,  $p < 0.01$ ) of the yeast cells expressing *lia<sup>2</sup>* (Fig. 3). These data suggested

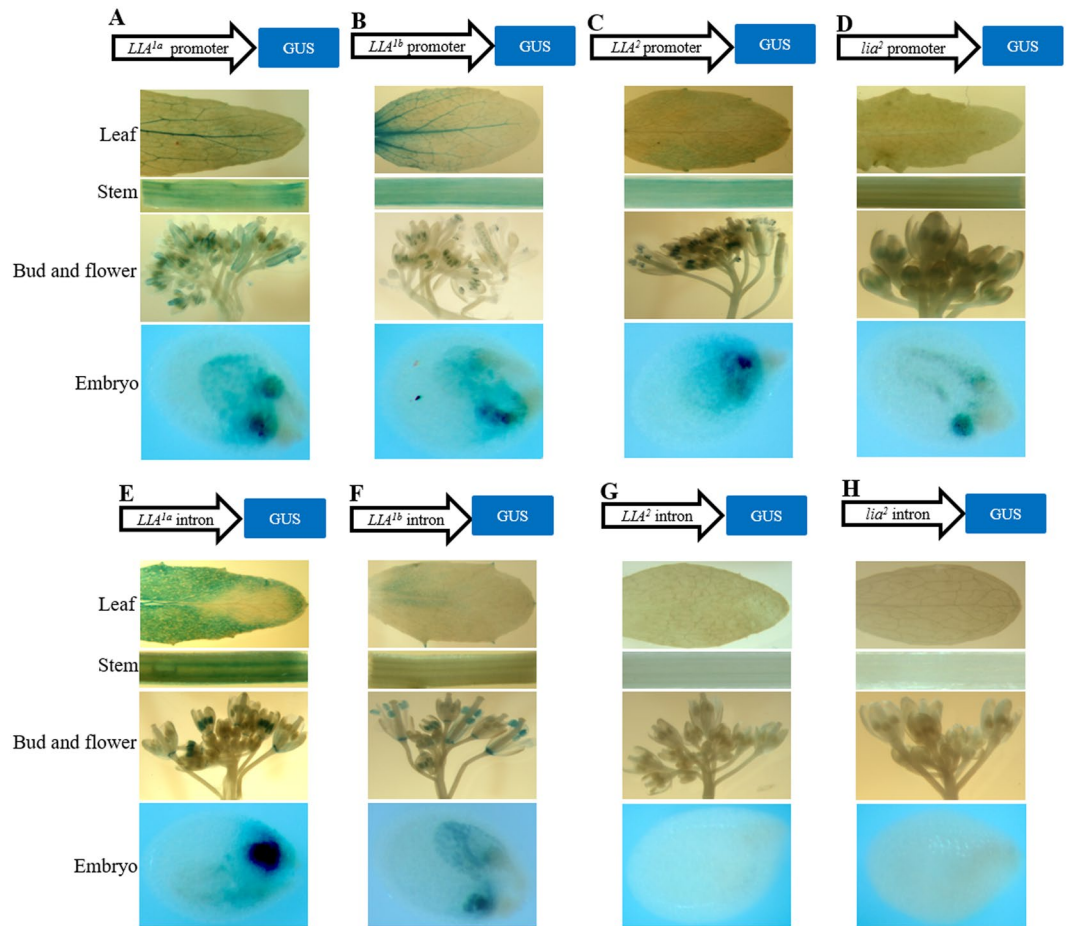


**Figure 4.** Expression analysis of the *SalFAD2.LIA1* alleles *LIA1<sup>a</sup>*, *LIA1<sup>b</sup>* and *lia1*, and *SalFAD2.LIA2* alleles *LIA2* and *lia2*. Actin 2 (FG576123) was used as an internal control. A. *LIA1<sup>a</sup>*, *LIA1<sup>b</sup>*, *lia1*, *LIA2* and *lia2* were expressed in the stem, leaf, flower, bud and 18d old embryo as revealed by RT-PCR. B. Quantitative Real-time PCR analysis of expression levels of the *LIA1<sup>a</sup>*, *LIA1<sup>b</sup>*, *lia1*, *LIA2* and *lia2* in 18d- old embryos of Y1798 (low), Y514 (medium) and Y1801 (high). Error bars indicated SD of the mean. \*\*Statistical significant difference at  $p = 0.01$  level.

that the cloned *SalFAD2.LIA2* alleles *LIA2* and *lia2* are functional, but the enzyme encoded by *lia2* produced lower C16:2 and C18:2 contents than that by *LIA2*. The C16:2 ( $2.5\% \pm 0.04\%$  SD) and C18:2 ( $9.7\% \pm 0.3\%$  SD) contents of the yeast culture expressing *LIA1<sup>a</sup>* were significantly lower than the C16:2 content ( $6.5\%$ ) ( $t = 3.95$ ,  $p < 0.01$ ) and C18:2 content ( $12.6\%$ ) ( $t = 3.36$ ,  $p < 0.01$ ) of the yeast cells expressing *LIA2* (Fig. 3), implying that the enzyme encoded by *LIA2* produced higher C16:2 and C18:2 contents than that by *LIA1<sup>a</sup>*. In summary, the cloned *SalFAD2.LIA1* alleles *LIA1<sup>a</sup>*, *LIA1<sup>b</sup>* and *lia1* and the *SalFAD2.LIA2* alleles *LIA2* and *lia2* encoded functional palmitoleate and oleate desaturases that were capable of converting C16:1 to C16:2 and C18:1 to C18:2, but control different C16:2 and C18:2 contents.

**Transcription analysis of the *SalFAD2.LIA1* and *SalFAD2.LIA2* alleles.** Transcripts of the *SalFAD2.LIA1* and *SalFAD2.LIA2* genes were detected in stem, leaf, flower bud and 18d-old embryo of Y1801, Y514 and Y1798 (Fig. 4A), suggesting that *LIA1<sup>a</sup>*, *LIA1<sup>b</sup>*, *lia1*, *LIA2* and *lia2* were constitutively expressed in the plant. The RT-PCR bands from *LIA1<sup>a</sup>*, *LIA1<sup>b</sup>*, *lia1*, *LIA2* and *lia2* (Fig. 4A) were cloned and sequenced, confirming the presence of a single copy. *LIA1<sup>a</sup>*, *LIA1<sup>b</sup>*, *lia1*, *LIA2* and *lia2* exhibited differences in the transcription level in the 18d-old embryo. *LIA1<sup>a</sup>* in Y1801 had the highest transcription level (Fig. 4B), and the transcription levels of *LIA1<sup>b</sup>* in Y514 and *lia1* in Y1798 were 12.3% (SD 0.1%) and 5.0% (SD 0.1%) of that of *LIA1<sup>a</sup>*, respectively. *lia1* had the lowest transcription level, at 40.8% (SD 0.1%) of *LIA1<sup>b</sup>*. *LIA2* in Y1801 and *LIA2* in Y514 had the same transcription level. *lia2* in Y1798 had a very low transcription level, at 1.1% (SD 0.1%) of *LIA2* (Fig. 4B). The transcription levels of *LIA1<sup>a</sup>*, *LIA1<sup>b</sup>*, *lia1*, *LIA2* and *lia2* were correlated with the C18:2 contents in the seeds of Y1801, Y514 and Y1798.

**Functional analysis of the putative promoters of the *SalFAD2.LIA1* and *SalFAD2.LIA2* alleles.** The constructs pBI101-*LIA1<sup>a</sup>*-promoter-GUS, pBI101-*LIA1<sup>b</sup>*-promoter-GUS, pBI101-*LIA2*-promoter-GUS and pBI101-*lia2*-promoter-GUS, carrying the putative promoter fragment of *LIA1<sup>a</sup>*, *LIA1<sup>b</sup>*, *LIA2* and *lia2*, respectively, were transformed into *Arabidopsis*. Histochemical assays led to the detection of GUS activity in the leaf, stem, bud, flower tissues and embryo (Fig. 5A–D) in the transgenic plants containing the constructs pBI101-*LIA1<sup>a</sup>*-Promoter-GUS, pBI101-*LIA1<sup>b</sup>*-Promoter-GUS, pBI101-*LIA2*-Promoter-GUS and pBI101-*lia2*-Promoter-GUS. These results indicated the putative promoter fragments of *LIA1<sup>a</sup>*, *LIA1<sup>b</sup>*, *LIA2* and

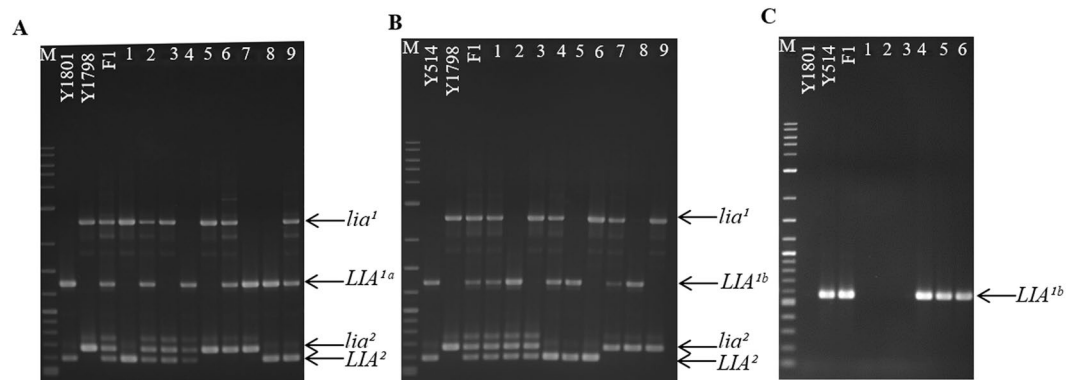


**Figure 5.** Histochemical localization of GUS activity in the leaf, stem, bud and flower, and 7d old embryo of transgenic *Arabidopsis* plants containing the construct pBI101-*LIA*<sup>1a</sup>-promoter-GUS with the promoter of *LIA*<sup>1a</sup> (A), pBI101-*LIA*<sup>1b</sup>-promoter-GUS with the promoter of *LIA*<sup>1b</sup> (B), pBI101-*LIA*<sup>2</sup>-promoter-GUS with the promoter of *LIA*<sup>2</sup> (C), pBI101-*lia*<sup>2</sup>-promoter-GUS with the promoter of *lia*<sup>2</sup> (D), pBI101-*LIA*<sup>1a</sup>-intron-GUS with the intron of *LIA*<sup>1a</sup> (E), pBI101-*LIA*<sup>1b</sup>-intron-GUS with the intron of *LIA*<sup>1b</sup> (F), pBI101-*LIA*<sup>2</sup>-intron-GUS with the intron of *LIA*<sup>2</sup> (G) and pBI101-*lia*<sup>2</sup>-intron-GUS with the intron of *lia*<sup>2</sup> (H), respectively. GUS staining was observed in the figures A, B, C, D, E and F, but not in the figures G and H.

*lia*<sup>2</sup> functioned as a constitutive promoter. Based on the extent of GUS staining, the promoters of *LIA*<sup>1a</sup> and *LIA*<sup>1b</sup> appeared to have similar level of expression, while the promoter of *LIA*<sup>2</sup> had a higher activity than that of *lia*<sup>2</sup>.

**The introns of the *SalFAD2.LIA1* alleles *LIA*<sup>1a</sup> and *LIA*<sup>1b</sup> function as a constitutive promoter.** The constructs pBI101-*LIA*<sup>1a</sup>-intron-GUS, pBI101-*LIA*<sup>1b</sup>-intron-GUS, pBI101-*LIA*<sup>2</sup>-intron-GUS and pBI101-*lia*<sup>2</sup>-intron-GUS containing the intron fragment of *LIA*<sup>1a</sup>, *LIA*<sup>1b</sup>, *LIA*<sup>2</sup> and *lia*<sup>2</sup>, respectively, were transformed into *Arabidopsis*. The intron of *lia*<sup>1</sup> was not tested for promoter function as it had a 1223 bp insertion at position -26. GUS activity was detected in the leaf, stem, bud, flower and the embryo of the transgenic plants containing the constructs pBI101-*LIA*<sup>1a</sup>-intron-GUS (Fig. 5E), pBI101-*LIA*<sup>1b</sup>-intron-GUS (Fig. 5F), indicating that the introns of *LIA*<sup>1a</sup> and *LIA*<sup>1b</sup> had the function of a constitutive promoter. However, the intron of *LIA*<sup>1b</sup> had a weaker promoter activity than that of *LIA*<sup>1a</sup> based on the extent of GUS staining. GUS activity was not detected in the transgenic plants containing the constructs of pBI101-*LIA*<sup>2</sup>-intron-GUS (Fig. 5G) and pBI101-*lia*<sup>2</sup>-intron-GUS (Fig. 5H), suggesting that the introns of *LIA*<sup>2</sup> and *lia*<sup>2</sup> had no promoter function.

**Co-segregation of the *SalFAD2.LIA1* and *SalFAD2.LIA2* allele-specific markers with C18:2 contents in the F<sub>2</sub> populations.** All of the nine possible genotypes were identified using the markers specific for *LIA*<sup>1a</sup>, *lia*<sup>1</sup>, *LIA*<sup>2</sup> and *lia*<sup>2</sup> in the cross of Y1798 (low) × Y1801 (high) (Fig. 6A). The homozygous F<sub>2</sub> plants of *LIA*<sup>1a</sup>*LIA*<sup>1a</sup>*LIA*<sup>2</sup>*LIA*<sup>2</sup> had a significantly higher C18:2 content (average: 21.2%) than the heterozygous F<sub>2</sub> plants of *LIA*<sup>1a</sup>*lia*<sup>1</sup>*LIA*<sup>2</sup>*LIA*<sup>2</sup> (average: 16.7%) ( $t = 3.08, p < 0.01$ ) and that of *LIA*<sup>1a</sup>*LIA*<sup>1a</sup>*LIA*<sup>2</sup>*lia*<sup>2</sup> (average: 17.9%) ( $t = 2.68, p < 0.01$ ). This result suggested that the allelic effect on the increase of C18:2 was *LIA*<sup>1a</sup> > *lia*<sup>1</sup> and *LIA*<sup>2</sup> > *lia*<sup>2</sup>. The homozygous F<sub>2</sub> plants of *LIA*<sup>1a</sup>*LIA*<sup>1a</sup>*lia*<sup>2</sup>*lia*<sup>2</sup> had a higher average C18:2 content (17.9%) than those of *lia*<sup>1</sup>*lia*<sup>1</sup>*LIA*<sup>2</sup>*LIA*<sup>2</sup> (average: 9.4%) ( $t = 7.80, p < 0.01$ ), indicating that *LIA*<sup>1a</sup> controlled a higher C18:2 content than *LIA*<sup>2</sup>. In the cross of Y1798 (low) × Y514 (medium), the nine genotypes for C18:2 content were distinguished from each other by the markers specific for *LIA*<sup>1b</sup>, *lia*<sup>1</sup>, *LIA*<sup>2</sup> and *lia*<sup>2</sup> (Fig. 6B). The average C18:2 content of the homozygous



**Figure 6.** Identification of different C18:2 genotypes based on the *SalFAD2.LIA1* and *SalFAD2.LIA2* allele-specific markers in the F<sub>2</sub> populations of the three crosses: (A) Y1798 (low) × Y1801 (high); B. Y1798 (low) × Y514 (medium) and C. Y1801 (high) × Y514 (medium). A. Y1801: high C18:2 line ( $LIA^{1a}LIA^{1a}LIA^2LIA^2$ ); Y1798: low C18:2 line ( $lia^1lia^1lia^2lia^2$ ); F<sub>1</sub> (Y1798 × Y1801):  $LIA^{1a}lia^1LIA^2lia^2$ ; Lane 1:  $lia^1lia^1LIA^2LIA^2$ ; Lane 2:  $LIA^{1a}lia^1LIA^2lia^2$ ; Lane 3:  $lia^1lia^1LIA^2lia^2$ . Lane 4:  $LIA^{1a}LIA^{1a}LIA^2lia^2$ ; Lane 5:  $lia^1lia^1lia^2lia^2$ ; Lane 6:  $LIA^{1a}lia^1lia^2lia^2$ ; Lane 7:  $LIA^{1a}LIA^{1a}lia^2lia^2$ ; Lane 8:  $LIA^{1a}LIA^{1a}LIA^2LIA^2$ ; Lane 9:  $LIA^{1a}lia^1LIA^2LIA^2$ . (B) Y514: medium C18:2 line ( $LIA^{1b}LIA^{1b}LIA^2LIA^2$ ); Y1798: low C18:2 line ( $lia^1lia^1lia^2lia^2$ ); F<sub>1</sub> (Y1798 × Y514):  $LIA^{1b}lia^1LIA^2lia^2$ ; Lane 1:  $LIA^{1b}lia^1LIA^2lia^2$ ; Lane 2:  $LIA^{1b}LIA^{1b}LIA^2lia^2$ ; Lane 3:  $lia^1lia^1LIA^2lia^2$ . Lane 4:  $LIA^{1b}lia^1LIA^2LIA^2$ ; Lane 5:  $LIA^{1b}LIA^{1b}LIA^2LIA^2$ ; Lane 6:  $lia^1lia^1LIA^2LIA^2$ ; Lane 7:  $LIA^{1b}lia^1lia^2lia^2$ ; Lane 8:  $LIA^{1b}LIA^{1b}lia^2lia^2$ ; Lane 9:  $lia^1lia^1lia^2lia^2$ . (C) Y1801: high C18:2 line ( $LIA^{1a}LIA^{1a}LIA^2LIA^2$ ); Y514: medium C18:2 line ( $LIA^{1b}LIA^{1b}LIA^2LIA^2$ ); F<sub>1</sub> (Y514 × Y1801):  $LIA^{1a}LIA^{1b}LIA^2LIA^2$ . 1–3:  $LIA^{1a}LIA^{1a}LIA^2LIA^2$ ; 4–6:  $LIA^{1b}LIA^2LIA^2$ .

F<sub>2</sub> plants ( $LIA^{1b}LIA^{1b}LIA^2LIA^2$ ) was 11.5%, which was significantly higher than that (average: 7.9%) of the heterozygous F<sub>2</sub> plants of  $LIA^{1b}lia^1LIA^2LIA^2$  ( $t = 3.95$ ,  $p < 0.01$ ), implying that  $LIA^{1b}$  contributed to a higher C18:2 content than  $lia^1$ .

In the cross of Y1801 (high) × Y514 (medium), the dominant marker specific for  $LIA^{1b}$  was used to identify the F<sub>2</sub> plants of  $LIA^{1a}LIA^{1a}LIA^2LIA^2$  from those of  $LIA^{1a}LIA^{1b}LIA^2LIA^2$  and  $LIA^{1b}LIA^{1b}LIA^2LIA^2$  (Fig. 6C). The homozygous F<sub>2</sub> plants of  $LIA^{1a}LIA^{1a}LIA^2LIA^2$  had an average C18:2 content of 23.7%, which was higher than the bulked F<sub>2</sub> plants of  $LIA^{1a}LIA^{1b}LIA^2LIA^2$  and  $LIA^{1b}LIA^{1b}LIA^2LIA^2$  (average: 21.9%) ( $t = 2.20$ ,  $p = 0.02$ ), suggesting that the allelic effect in increasing C18:2 content is  $LIA^{1a} > LIA^{1b}$ . The *SalFAD2.LIA1* and *SalFAD2.LIA2* genes co-localized with the QTL peaks on Sal01 and Sal08, respectively (Fig. 1).

## Discussion

In the present study, QTL mapping revealed the occurrence of four C18:2 QTLs in yellow mustard. The C18:2 QTLs on Sal01 and Sal08 were responsible for 14.9–47.3% and 9.5% of the phenotypic variation and co-localized with the *FAD2* genes *SalFAD2.LIA1* and *SalFAD2.LIA2*, respectively. However, the C18:2 QTL on Sal02 was mapped to the same region as the C18:3 QTL and *FAD3* gene *SALFAD3.LA1*, and the C18:2 QTL on Sal03 as the C22:1 QTL and the *FAE1* gene. The co-localization of the C18:2 QTL with the C18:3 and C22:1 QTLs could be due to the fact that the three fatty acids share the same biosynthetic pathway. Both C18:2 and C22:1 use C18:1n-9 as a starting substrate. The use of this substrate for C22:1 synthesis may result in less C18:1n-9 being available for C18:2 productions. Further desaturation of C18:2 by the *FAD3* gene leads to the production of C18:3 and the reduction of the C18:2 content.

The *SalFAD2.LIA1* and *SalFAD2.LIA2* genes each contained two exons, one intron within the 5'-UTR and a promoter adjacent to the first exon, which is in agreement with that in *Arabidopsis*<sup>4</sup>, *B. napus*<sup>5,6,32</sup> and other species including *G. hirsutum*<sup>7</sup> and *V. labrusca*<sup>10</sup>. The promoters and introns of *SalFAD2.LIA1* gene in yellow mustard demonstrated constitutive promoter activity as that of the *FAD2* in *B. napus*<sup>6</sup>. The protein encoded by each of the *SalFAD2.LIA1* alleles  $LIA^{1a}$ ,  $LIA^{1b}$  and  $lia^1$  and *SalFAD2.LIA2* alleles  $LIA^2$  and  $lia^2$  has six transmembrane domains and three H boxes harboring eight iron-binding histidines, which are consistent with previous reports in *Arabidopsis*<sup>4,12</sup>. The positions of the H boxes and histidines in the *FAD2* polypeptides in yellow mustard are in agreement with that in *Arabidopsis* (Supplementary Fig. S5). Heterologous expression in yeast confirmed that the cloned *SalFAD2.LIA1* alleles  $LIA^{1a}$  and  $LIA^{1b}$  ( $lia^1$ ) and the *SalFAD2.LIA2* alleles  $LIA^2$  and  $lia^2$  encoded functional palmitoleate and oleate desaturases that were capable of converting C16:1 to C16:2 and C18:1 to C18:2.

It has been reported that SNP/indel mutations in the coding region of the *FAD2* resulted in loss-of-function alleles in *B. napus*<sup>14–16</sup> and *B. rapa*<sup>17</sup>. However, the mutations observed in the coding region of *SalFAD2.LIA1* and *SalFAD2.LIA2* genes in yellow mustard were different from those reported in *Brassica* species<sup>14–17</sup>. Moreover, this report reveals that extensive SNP and indel mutations in the intron and promoter as well as transposable element insertion in the intron are involved in modulating the expression of *FAD2* in yellow mustard. The dominant *SalFAD2.LIA1* allele  $LIA^{1a}$  conditioned a higher C18:2 content and had a promoter with stronger activity than the dominant *SalFAD2.LIA2* allele  $LIA^2$ . In addition, the intron of  $LIA^{1a}$  demonstrated strong promoter activity, whereas that of  $LIA^2$  lost the promoter function. Sequence alignment with  $LIA^{1a}$  indicated that  $LIA^2$  harboured 103 SNPs and 29 indels in the promoter, and 106 SNPs and 44 indels in the intron. Search for cis-regulatory



elements in the promoters and introns of *LIA*<sup>1a</sup> and *LIA*<sup>2</sup> were performed using PlantCARE (<http://bioinformatics.psb.ugent.be/webtools/plantcare/html/>)<sup>33</sup>. The promoter of *LIA*<sup>1a</sup> contained 14 “CAAT box”, a common cis-acting element in the promoter and enhancer regions; while that of *LIA*<sup>2</sup> had only 5 “CAAT box”. The reduction of the number of “CAAT box” might account for the reduced promoter activity of *LIA*<sup>2</sup>. The intron of *LIA*<sup>1a</sup> contained 24 “CAAT box”. However, the intron of *LIA*<sup>2</sup> had only 9 “CAAT box”. Moreover, based on the prediction of possible transcription promoter using BDGP ([http://www.fruitfly.org/seq\\_tools/promoter.html](http://www.fruitfly.org/seq_tools/promoter.html))<sup>34</sup>, the intron of *LIA*<sup>1a</sup> contained the predicted transcription starting site “T” at nucleotide 825 and putative “TATA-box” TTTAAAA at position 787 (Supplementary Fig. S12). However, the intron of *LIA*<sup>2</sup> did not have the possible transcription promoter, which might have led to the loss of the promoter function. Therefore, it could be inferred that the mutations in the promoter and intron of *LIA*<sup>2</sup> resulted in reduced transcription level and C18:2 content.

The *SalFAD2.LIA1* allele *LIA*<sup>1b</sup> differed from *LIA*<sup>1a</sup> in the CDS, promoter and intron, and controlled a lower C18:2 content than *LIA*<sup>1a</sup>. Expression studies in *S. cerevisiae* indicated that *LIA*<sup>1a</sup> and *LIA*<sup>1b</sup> produced similar levels of C16:2 and C18:2 fatty acids although the proteins encoded by *LIA*<sup>1b</sup> harboured the amino acid changes at positions 72 and 84. Sequence alignment with *LIA*<sup>1a</sup> indicated that the promoter of *LIA*<sup>1b</sup> carried 9 SNP and 8 indel mutations. However, based on GUS staining, the promoters of *LIA*<sup>1a</sup> and *LIA*<sup>1b</sup> appeared to have similar level of expression. Therefore, the mutations of the CDS and promoter of *LIA*<sup>1b</sup> had no effect on the biosynthesis of C18:2. Sixteen indels and 49 SNP mutations occurred in the intron of *LIA*<sup>1b</sup> compared with that of *LIA*<sup>1a</sup>. Searching for cis-regulatory elements indicated that the intron of *LIA*<sup>1a</sup> contains the motif “5UTR Py-rich stretch” (TTTCTTCTTT) (nucleotides 856–866) (Supplementary Fig. S9) which can enhance transcription level. However, the intron of *LIA*<sup>1b</sup> had lost this motif due to point mutations at positions 868, 871 and 872 (Supplementary Fig. S9). This might lead to the reduction of the promoter activity of the intron in *LIA*<sup>1b</sup>. Thus, it could be concluded that the mutation in the intron of *LIA*<sup>1b</sup> resulted in the reduction of C18:2 content. The recessive *SalFAD2.LIA1* allele *lia*<sup>1</sup> produced a lower C18:2 content than *LIA*<sup>1b</sup>. The allele *lia*<sup>1</sup> was identical to *LIA*<sup>1b</sup> in the promoter, exon 1 and exon 2, but harboured a 1223 bp transposable element insertion at position –26 bp in the 5' upstream region from the translation start codon. The insertion of the retrotransposon in the intron of *LIA*<sup>1b</sup> might have generated the allele *lia*<sup>1</sup> with further reduced transcription level and C18:2 content. There are two possible explanations for this: Firstly, the inserted retrotransposon of *lia*<sup>1</sup> might disturb the interaction of the promoter, intron and other transcriptional factors, thereby affecting *FAD2* expression. Secondly, the promoter of *lia*<sup>1</sup> is further away from the CDS due to the insertion, which might result in a reduction in transcription efficiency compared to *LIA*<sup>1b</sup>, where the regulatory region is adjacent to the coding region. Insertions of DNA transposons or retrotransposons within or near genes have been reported to negatively affect the expression of genes by decreasing or abolishing transcription<sup>35–39</sup>.

The recessive *SalFAD2.LIA2* allele *lia*<sup>2</sup> controlled a lower C18:2 content than *LIA*<sup>2</sup>, which could be due to the mutations of the CDS and promoter since the introns of both *LIA*<sup>2</sup> and *lia*<sup>2</sup> lost the promoter function. Yeast cells expressing *lia*<sup>2</sup> had significantly lower C16:2 and C18:2 contents than the yeast cells expressing *LIA*<sup>2</sup>. Compared with *LIA*<sup>2</sup>, the polypeptide encoded by *lia*<sup>2</sup> comprised six amino acid changes at positions 12, 42, 58, 63, 169 and 206 (Supplementary Fig. S5), which might result in reduced enzyme activity. GUS staining indicated that the promoter of *lia*<sup>2</sup> was weaker and had a lower expression level than that of *LIA*<sup>2</sup>. Sequence alignment with *LIA*<sup>2</sup> revealed that the promoter of *lia*<sup>2</sup> harboured 9 indels. The predicted site of “TATA-box” in the promoter region of *lia*<sup>2</sup> was different from that of *LIA*<sup>2</sup> due to a 10 bp insertion occurred at position 368 (Supplementary Fig. S11). This might have resulted in the reduced promoter function of *lia*<sup>2</sup>. Therefore, it was concluded that *lia*<sup>2</sup> produced a low C18:2 content due to the mutations of both CDS and promoter.

In conclusion, this study revealed that the linoleic acid variation was controlled by multiple alleles with different expression levels at two gene loci, *SalFAD2.LIA1* and *SalFAD2.LIA2*, in yellow mustard. The results demonstrated that complex quantitative genetic variation of trait phenotype could be caused by multiple alleles of oligogenic loci resulting from mutations in regulatory regions such as promoter and intron as well as CDS.

## Materials and Methods

**Plant materials.** Y1798 and Y1801, developed via pedigree breeding from the variety Andante, are S<sub>4</sub> and S<sub>5</sub> inbred lines, respectively. Y514 is the doubled haploid line SaMD3<sup>30</sup>. Linoleic, linolenic (C18:3) and erucic (C22:1) acid contents of the three parental lines Y1798, Y514 and Y1801 are shown in Table 1. Y1798 has a low C18:2 content (average: 4.2%; range: 3.5–5.3%), an average C18:3 content of 13.0% (range: 10.9–15.5%) and an average C22:1 content of 36.3% (range: 11.2–45.6%). Y514 has a medium C18:2 content (average: 12.5%; range: 12.0–13.0%), an average C18:3 content of 10.8% (range: 9.6–12.4%) and zero C22:1 content. Y1801 has a high C18:2 content (average: 31.5%; range: 28.7–37.7%), an average C18:3 content of 12.5% (range: 10.2–14.5%) and an average C22:1 content of 16.1% (range: 14.7–17.3%). The C18:3 and C22:1 genotypes of Y1798, Y514 and Y1801 were determined using the two *FAD3* genes, *SalFAD3.LA1* and *SalFAD3.LA2*, and *FAE1* gene allele-specific markers<sup>27, 40</sup> (Supplementary Table S3). Y1798 carried the *FAE1* allele *E*<sup>1</sup> for a high C22:1 content, and *SalFAD3.LA1* allele *LA*<sup>1a</sup> and *SalFAD3.LA2* allele *LA*<sup>2</sup> for a high C18:3 content. Y514 had the *FAE1* allele *e* for zero C22:1 content, and *SalFAD3.LA1* allele *LA*<sup>1</sup> for a high C18:3 content and *SalFAD3.LA2* allele *la*<sup>2</sup> for a low C18:3 content. Y1801 harboured the *FAE1* allele *E*<sup>2</sup> for a medium C22:1 content and the *SalFAD3.LA1* allele *la*<sup>1</sup> for a low C18:3 content and *SalFAD3.LA2* allele *LA*<sup>2</sup> for a high C18:3 content (Supplementary Table S3). The F<sub>1</sub> seeds of the three crosses Y1798 (low) × Y1801 (high), Y1798 (low) × Y514 (medium) and Y1801 (high) × Y514 (medium) were produced. The F<sub>1</sub> plants were self-pollinated to produce F<sub>2</sub> seeds. One hundred and twenty-three F<sub>2</sub> plants from one F<sub>1</sub> plants of Y1798 × Y1801, 122 F<sub>2</sub> plants from one F<sub>1</sub> plant of Y1798 × Y514 and 122 F<sub>2</sub> plants from one F<sub>1</sub> plant of Y1801 × Y514 were used for constructing the genetic linkage map and QTL mapping of C18:2, C18:3 and C22:1 contents. All plants were raised under the same conditions in the greenhouse at Agriculture and Agri-Food Canada-Saskatoon Research and Development Centre.

**Seed fatty acid analysis.** Seed fatty acid composition was analyzed according to Thies (1971)<sup>41</sup> with the following modification: the gas chromatography of the methyl esters was performed with a HP-INNOWax fused silica capillary column (0.25 mm by 0.5 m and 7.5  $\mu$ m) (Agilent Technologies) at 250 °C using hydrogen as the carrier gas. A minimum of 10 seeds from each of the parental lines and F<sub>1</sub> hybrids, and 150 F<sub>2</sub> seeds of each of the crosses Y1798  $\times$  Y1801, Y1798  $\times$  Y514 and Y1801  $\times$  Y514 were half-seed analyzed according to Downey and Harvey (1963)<sup>42</sup>.

**Cloning of the coding regions of the *SalFAD2.LIA1* alleles *LIA*<sup>1a</sup>, *LIA*<sup>1b</sup> and *lia*<sup>1</sup>, and *SalFAD2.LIA2* alleles *LIA*<sup>2</sup> and *lia*<sup>2</sup>.** Primer pair No 1 (Supplementary Table S1) was designed based on the conserved coding regions of *BnFAD2* (FJ907397) from *B. napus* and *AtFAD2* (TAIR: At3G12120) from *A. thaliana*, and used to clone the CDS of the *FAD2* alleles *LIA*<sup>1a</sup>, *LIA*<sup>1b</sup>, *lia*<sup>1</sup>, *LIA*<sup>2</sup> and *lia*<sup>2</sup> in yellow mustard. Genomic DNA of each of the three parental lines Y1798, Y514 and Y1801 was used as template for PCR amplification with *fidelity-Taq* DNA polymerase (Affymetrix) in a thermocycler with 30 cycles of the following program: 94 °C for 30 s, 56 °C for 30 s, and 68 °C for 2 min.

**Cloning the 3' and 5' flanking sequences of the coding regions of the *SalFAD2.LIA1* alleles *LIA*<sup>1a</sup>, *LIA*<sup>1b</sup> and *lia*<sup>1</sup>, and *SalFAD2.LIA2* alleles *LIA*<sup>2</sup> and *lia*<sup>2</sup>.** The primer pairs used to clone the 3' and 5' flanking sequences of the coding regions of the *SalFAD2.LIA1* alleles *LIA*<sup>1a</sup>, *LIA*<sup>1b</sup> and *lia*<sup>1</sup>, and *SalFAD2.LIA2* alleles *LIA*<sup>2</sup> and *lia*<sup>2</sup> are presented in Supplementary Table S1 and Supplementary Fig. S4. Primer pair No 2 was designed based on the sequence of the 3' coding region of the *FAD2* alleles *LIA*<sup>1a</sup>, *LIA*<sup>1b</sup>, *lia*<sup>1</sup>, *LIA*<sup>2</sup> and *lia*<sup>2</sup> and used to clone the 3' downstream sequence of each of the five alleles. Primer pair No 3, designed based on the 5' coding region of *LIA*<sup>1a</sup>, *LIA*<sup>1b</sup> and *lia*<sup>1</sup>, was used to clone the 5' upstream sequences of the three alleles. Primer pair No 4 were designed based on the 5' upstream sequence obtained from the first round of PCR walking to clone the further upstream sequences of *LIA*<sup>1a</sup>, *LIA*<sup>1b</sup> and *lia*<sup>1</sup>. Primer pair No 5, designed based on the promoter sequence of *BnFAD2* (KF038144), was used to clone the promoter region of *LIA*<sup>1a</sup>, *LIA*<sup>1b</sup> and *lia*<sup>1</sup>. Primer pair No 6 was designed to clone the DNA fragment between the promoter and 5' flanking sequence of *LIA*<sup>1a</sup>, *LIA*<sup>1b</sup> and *lia*<sup>1</sup>. Primer pair No 7, designed based on the 5' coding region was used to clone the 5' upstream sequences of *LIA*<sup>2</sup> and *lia*<sup>2</sup>. Primer pair No 8, designed based on the 5' upstream sequence obtained from the first round of PCR walking to clone the further upstream sequences of *LIA*<sup>2</sup> and *lia*<sup>2</sup>. PCR walking was performed according to the protocol of Siebert *et al.* (1995)<sup>43</sup>. The standard protocol from the clontech kit (Protocol PT 3042, Version PR 03300) was followed to facilitate the PCR walking. Primer pair No 9 was designed to amplify *LIA*<sup>1a</sup>, *LIA*<sup>1b</sup> and *lia*<sup>1</sup> and primer pair No 10 was designed to amplify *LIA*<sup>2</sup> and *lia*<sup>2</sup> in their entirety. The sequence analysis of the inserted DNA fragment in the intron of the *SalFAD2.LIA1* allele *lia*<sup>1</sup> was according to Zeng and Cheng (2014)<sup>27</sup>. Primer pair No 11 (Supplementary Table S1) was designed based on the exon 1 (nucleotide: 1 to 20) and exon 2 (nucleotide: 1156–1137) of *LIA*<sup>1a</sup> to obtain the cDNA sequence of the *SalFAD2.LIA1* alleles *LIA*<sup>1a</sup>, *LIA*<sup>1b</sup>, *lia*<sup>1</sup>. Primer pair No 12 (Supplementary Table S1) was designed based on the exon 1 (nucleotide: 1 to 20) and exon 2 (nucleotide: 1162–1143) of *LIA*<sup>2</sup> to get the cDNA sequence of the *SalFAD2.LIA2* alleles *LIA*<sup>2</sup> and *lia*<sup>2</sup>.

**Transformation of yeast and fatty acid analysis.** The coding sequences of the *SalFAD2.LIA1* alleles *LIA*<sup>1a</sup> and *LIA*<sup>1b</sup> (*lia*<sup>1</sup>) were amplified using primer pair No 13 and those of the *SalFAD2.LIA2* alleles *LIA*<sup>2</sup> and *lia*<sup>2</sup> using primer pair No 14 (Supplementary Table S1). The amplified sequences of *LIA*<sup>1a</sup>, *LIA*<sup>1b</sup> (*lia*<sup>1</sup>), *LIA*<sup>2</sup> and *lia*<sup>2</sup> were then cloned into the pYES2.1/V5-His-TOPO expression vector (Invitrogen), respectively, and sequenced to confirm the correct orientation of genes. The four constructs pYES2.1/V5-His-TOPO-*LIA*<sup>1a</sup>, pYES2.1/V5-His-TOPO-*LIA*<sup>1b</sup> (*lia*<sup>1</sup>), pYES2.1/V5-His-TOPO-*LIA*<sup>2</sup> and pYES2.1/V5-His-TOPO-*lia*<sup>2</sup> were transformed into *Saccharomyces cerevisiae* strain Inv Sc1 (Invitrogen), by the lithium acetate method<sup>44</sup>. Yeast cells transformed with the empty vector pYES2.1/V5-His-TOPO plasmid were used as a control. Transgenic cells were screened in complete minimal drop-out uracil medium containing 2% raffinose as a carbon source at 30 °C. The positive clones were grown at 30 °C overnight in minimal media supplemented with 2% raffinose and lacking uracil. The transgenic cell culture was centrifuged, followed by washing, and then used to inoculate 20 ml induction media (minimal media lacking uracil and supplemented with 2% galactose and 1% raffinose) to an OD<sub>600</sub> of 0.5. Cultures were grown overnight. The fatty acid composition analysis was the same as previously reported<sup>27</sup>.

**RT-PCR and Quantitative Real-time PCR.** Total RNA was extracted from leaf, stem, flower bud, flower and 18-d old embryos of Y1798, Y514 and Y1801 using an RNeasy plant mini kit (Qiagen). Extracted RNA was treated with DNase I (Ambion), and cDNA was synthesized using the ReverTra Ace-a-First Strand cDNA synthesis kit (Thermo Fisher) according to the manufacturer's instruction.

Primer pair No 15 (Supplementary Table S1) was designed based on exon 1 and exon 2 sequences of the *SalFAD2.LIA1* alleles *LIA*<sup>1a</sup>, *LIA*<sup>1b</sup> and *lia*<sup>1</sup>, while primer pair No 16 (Supplementary Table S1) was designed based on exon 1 and exon 2 sequences of the *SalFAD2.LIA2* alleles of *LIA*<sup>2</sup> and *lia*<sup>2</sup>. To further detect the differences in transcription level, quantitative PCR (qPCR) analysis was conducted on 18d old embryos of Y1798, Y514 and Y1801. Primer pair No 17 (Supplementary Table S1), which was used to amplify the 155 bp fragment (nucleotides 11 to 165) of the cDNA, specific for the *SalFAD2.LIA1* alleles *LIA*<sup>1a</sup>, *LIA*<sup>1b</sup> and *lia*<sup>1</sup>. Primer pair No 18 (Supplementary Table S1), which was used to amplify the 191 bp fragments (nucleotides 106 to 296) of the cDNA, specific for the *SalFAD2.LIA2* alleles *LIA*<sup>2</sup> and *lia*<sup>2</sup>. The qPCR analysis was performed with SsoFast EvaGreen supermix (Bio-Rad) according to the manufacturer's instructions using a Bio-Rad CFX96™ system. Primer pair No 19 (Supplementary Table S1) specific for Actin2 (FG576123) was used as an internal control for normalization. Three separate first strand cDNA reactions were analyzed in duplicate for each sample, and expression levels were calculated as described by Livak and Schmittgen (2001)<sup>45</sup>.

**Transformation of *Arabidopsis* and histochemical GUS Assays.** The putative promoter sequences of 639 bp, 714 bp, 566 bp and 554 bp for *LIA*<sup>1a</sup>, *LIA*<sup>1b</sup>, *LIA*<sup>2</sup> and *lia*<sup>2</sup> were cloned using primer pairs No 20, 21, 22 and 23, respectively (Supplementary Table S1), and inserted into the plant expression vector pBI101 upstream of the GUS gene. The intron sequences of 1286 bp, 1354 bp, 579 bp and 617 bp for *LIA*<sup>1a</sup>, *LIA*<sup>1b</sup>, *LIA*<sup>2</sup> and *lia*<sup>2</sup> were cloned using primer pairs No 24, 25, 26 and 27, respectively (Supplementary Table S1) and were also inserted into the plant expression vector pBI101 upstream of the GUS gene. The resulting constructs were transformed into electrocompetent *Agrobacterium* cells GV3101 by electroporation according to the manufacturer's instructions. *Agrobacterium*-mediated transformation of wild-type *Arabidopsis* plants was performed according to Clough and Bent (1998)<sup>46</sup>. Transgenic plants were screened and analyzed according to Jako *et al.*<sup>47</sup>. For each construct, at least 10 positive plants were used for analysis. Embryos of 7d-old, leaf, stem, bud and flower tissues of the transgenic plants were stained overnight in solution containing 0.5 mg/ml 5-bromo-4-chloro-3-indolyl-beta-D-glucuronide, 50 mM Na phosphate buffer, pH 7.0, 1 mM K3Fe(CN)6, 1 mM K4Fe(CN)6, 20 mM EDTA. The staining of the embryo and vegetative tissues was observed and photographed using a Zeiss stereomicroscope with a color CCD camera.

**Development of the *SalFAD2.LIA1* and *SalFAD2.LIA2* allele-specific markers.** Primer pair No 28 (Supplementary Table S1), which was designed based on the conserved flanking sequences of the intron of the *SalFAD2.LIA1* and *SalFAD2.LIA2* genes and produced co-dominant markers of 1317 bp, 1329 bp, 2549 bp, 607 bp and 682 bp specific for the *SalFAD2.LIA1* alleles *LIA*<sup>1a</sup>, *LIA*<sup>1b</sup> and *lia*<sup>1</sup>, and *SalFAD2.LIA2* alleles *LIA*<sup>2</sup> and *lia*<sup>2</sup>, respectively. The PCR was performed using phusion high-fidelity polymerase (NEB) with 30 cycles of the following program: 98 °C for 10 s, 60 °C for 30 s, and 72 °C for 2 min. Primer pair No 29 (Supplementary Table S1), designed based on the 12 bp insertion in the promoter region of *LIA*<sup>1b</sup>, generating a dominant marker of 548 bp specific for *LIA*<sup>1b</sup>. The PCR amplification was the following program of 30 cycles: 94 °C for 30 s, 56 °C for 30 s, and 72 °C for 1 min.

**Construction of a linkage map and QTL analysis.** A total of 785 ILP primer pairs, 164 from *A. thaliana*, 316 from *B. napus* and 305 from *B. rapa*, were used to screen the parental lines Y1798 (low), Y514 (medium) and Y1801 (high) for polymorphic markers. Of the 785 primer pairs, 130, 110 and 114 polymorphic markers were generated between the parental lines Y1798 and Y1801, Y1798 and Y514, and Y1801 and Y514, respectively. The genetic linkage map was constructed by using JoinMap<sup>48</sup> version 4.0 at LOD scores  $\geq 4.0$ . Chi-square test for goodness-of-fit was performed to determine if marker segregation deviated from the expected ratio. The threshold of  $p < 0.01$  was used to exclude the distorted markers from the map construction. An MQM mapping analysis was conducted using the MapQTL 6.0 software<sup>49</sup> to detect QTLs for C18:2, C18:3 and C22:1 contents. Permutation test (1,000 replications) was used to determine the significance level for LOD with a genome-wide probability of  $p < 0.05$ . QTL analysis was performed for C18:2, linolenic (C18:3) and erucic (C22:1) contents in the F<sub>2</sub> populations derived from the three crosses Y1798 (low)  $\times$  Y1801 (high), Y1798 (low)  $\times$  Y514 (medium) and Y1801 (high)  $\times$  Y514 (medium). The effects of the *SalFAD2.LIA1* alleles *LIA*<sup>1a</sup>, *LIA*<sup>1b</sup> and *lia*<sup>1</sup>, and *SalFAD2.LIA2* alleles *LIA*<sup>2</sup> and *lia*<sup>2</sup> on linoleic (C18:2) acid content was estimated by ANOVA procedure.

**Accession numbers.** The GenBank accession numbers for the nucleotide sequences of the *SalFAD2.LIA1* alleles *LIA*<sup>1a</sup>, *LIA*<sup>1b</sup> and *lia*<sup>1</sup> are KY305533, KY305534, and KY305535 and *SalFAD2.LIA2* alleles *LIA*<sup>2</sup> and *lia*<sup>2</sup> are KY305536 and KY305537, respectively.

## References

- Márquez-Ruiz, G., Garcés, R., León-Camacho, M. & Mancha, M. Thermoxidative stability of triacylglycerols from mutant sunflower seeds. *J Am Oil Chem Soc* **76**, 1169–1174 (1999).
- Cunnane, S. C. Problems with essential fatty acids: time for a new paradigm? *Prog Lipid Res.* **42**, 544–568 (2003).
- Shanklin, J. & Cahoon, E. B. Desaturation and related modifications of fatty acids 1. *Annu. Rev. Plant Biol.* **49**, 611–641 (1998).
- Okuley, J. *et al.* *Arabidopsis FAD2* gene encodes the enzyme that is essential for polyunsaturated lipid synthesis. *Plant Cell* **6**, 147–158 (1994).
- Lee, K.-R. *et al.* Functional analysis and tissue-differential expression of four *FAD2* genes in amphidiploid *Brassica napus* derived from *Brassica rapa* and *Brassica oleracea*. *Gene* **531**, 253–262 (2013).
- Xiao, G. *et al.* Characterization of the promoter and 5'-UTR intron of oleic acid desaturase (*FAD2*) gene in *Brassica napus*. *Gene* **545**, 45–55 (2014).
- Zhang, D. *et al.* Identification and expression of a new delta-12 fatty acid desaturase (*FAD2-4*) gene in upland cotton and its functional expression in yeast and *Arabidopsis thaliana* plants. *Plant Physiol. Biochem.* **47**, 462–471 (2009).
- Jin, U.-H. *et al.* Characterization and temporal expression of a  $\omega$ -6 fatty acid desaturase cDNA from sesame (*Sesamum indicum* L.) seeds. *Plant Sci.* **161**, 935–941 (2001).
- Hernández, M. L., Mancha, M. & Martínez-Rivas, J. M. Molecular cloning and characterization of genes encoding two microsomal oleate desaturases (*FAD2*) from olive. *Phytochemistry* **66**, 1417–1426 (2005).
- Lee, K.-R. *et al.* Molecular cloning and functional analysis of two *FAD2* genes from American grape (*Vitis labrusca* L.). *Gene* **509**, 189–194 (2012).
- Kim, M. J. *et al.* Seed-specific expression of sesame microsomal oleic acid desaturase is controlled by combinatorial properties between negative cis-regulatory elements in the *SeFAD2* promoter and enhancers in the 5'-UTR intron. *Mol. Genet. Genomics* **276**, 351–368 (2006).
- Shanklin, J., Whittle, E. & Fox, B. G. Eight histidine residues are catalytically essential in a membrane-associated iron enzyme, stearoyl-CoA desaturase, and are conserved in alkane hydroxylase and xylene monooxygenase. *Biochemistry* **33**, 12787–12794 (1994).
- Kurdrid, P., Subudhi, S., Hongsthong, A., Ruengjitchawalya, M. & Tanticharoen, M. Functional expression of Spirulina- $\Delta 6$  desaturase gene in yeast. *Saccharomyces cerevisiae*. *Mol. Biol. Rep.* **32**, 215–226 (2005).
- Hu, X., Sullivan-Gilbert, M., Gupta, M. & Thompson, S. A. 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 (2006).
- Yang, Q. *et al.* Identification of *FAD2* and *FAD3* genes in *Brassica napus* genome and development of allele-specific markers for high oleic and low linolenic acid contents. *Theor. Appl. Genet.* **125**, 715–729 (2012).

16. Wells, R. *et al.* The control of seed oil polyunsaturate content in the polyploid crop species *Brassica napus*. *Mol. Breed.* **33**, 349–362 (2014).
17. Tanhuanpää, P., Vilkkki, J. & Vihinen, M. 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 (1998).
18. Warwick, S. I. & Black, L. D. Molecular systematics of *Brassica* and allied genera (subtribe Brassicinae, Brassicaceae)—chloroplast genome and cytodeme congruence. *Theor. Appl. Genet.* **82**, 81–92 (1991).
19. Drost, W.J., Rakow, G. & Raney, P. Inheritance of erucic acid content in yellow mustard (*Sinapis alba* L.). In Proceedings of the 10th International Rapeseed Congress, Canberra, Australia (1999a).
20. Drost, W.J., Rakow, G. & Raney, P. Inheritance of glucosinolate content in yellow mustard (*Sinapis alba* L.). In Proceeding of the 10th International Rapeseed Congress, Canberra, Australia (1999b).
21. Cheng, B., Williams, D. J. & Zhang, Y. Genetic variation in morphology, seed quality and self-(in) compatibility among the inbred lines developed from a population variety in outcrossing yellow mustard (*Sinapis alba*). *Plants* **1**, 16–26 (2012).
22. Han, J. X. *et al.* Functional characterization of  $\beta$ -ketoacyl-CoA synthase genes from *Brassica napus* L. *Plant Mol. Biol.* **46**, 229–239 (2001).
23. Roscoe, T., Lessire, R., Puyaubert, J., Renard, M. & Delseny, M. Mutations in the *fatty acid elongation 1* gene are associated with a loss of  $\beta$ -ketoacyl-CoA synthase activity in low erucic acid rapeseed. *FEBS Lett.* **492**, 107–111 (2001).
24. Katavic, V. *et al.* Restoring enzyme activity in nonfunctional low erucic acid *Brassica napus* fatty acid elongase 1 by a single amino acid substitution. *Eur J Biochem* **269**, 5625–5631 (2002).
25. Wu, Y., Xiao, L., Wu, G. & Lu, C. Cloning of *fatty acid elongase1* gene and molecular identification of A and C genome in *Brassica* species. *Sci China Ser C-Life Sci* **3**, 343–349 (2007).
26. Wu, G., Wu, Y. H., Xiao, L., Li, X. D. & Lu, C. M. 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 (2008).
27. Zeng, F. & Cheng, B. Transposable element insertion and epigenetic modification cause the multiallelic variation in the expression of *FAE1* in *Sinapis alba*. *Plant Cell* **26**, 2648–2659 (2014).
28. Yaniv, Z., Schafferman, D., Elber, Y., Ben-Moshe, E. & Zur, M. Evaluation of *Sinapis alba*, native to Israel, as a rich source of erucic acid in seed oil. *Ind Crops Prod.* **2**, 137–142 (1994).
29. Cheng, B., Javidfar, F. & Roslinsky, V. Development and genetic studies of inbred lines with different linoleic acid contents in yellow mustard. In Proceedings of the 14th International Rapeseed Congress, Saskatoon, Canada (2015).
30. Bundrock, T. Doubled haploidy in yellow mustard (*Sinapis alba* L.). Master thesis, University of Saskatchewan, Saskatoon, SK, Canada. (1998).
31. Javidfar, F. & Cheng, B. Single locus, multiallelic inheritance of erucic acid content and linkage mapping of *FAE1* gene in yellow mustard. *Crop Sci.* **53**, 825–832 (2013).
32. Jung, J. H. *et al.* Identification of functional *BrFAD2-1* gene encoding microsomal delta-12 fatty acid desaturase from *Brassica rapa* and development of *Brassica napus* containing high oleic acid contents. *Plant Cell Rep.* **30**, 1881–1892 (2011).
33. Lescot, M. *et al.* PlantCARE, a database of plant cis-acting regulatory elements and a portal to tools for in silico analysis of promoter sequences. *Nucleic Acids Res.* **30**(1), 325–327 (2002).
34. Reese, M. G. Application of a time-delay neural network to promoter annotation in the *Drosophila melanogaster* genome. *Comput. Chem.* **26**(1), 51–6 (2001).
35. Kumar, A. & Bennetzen, J. L. Plant retrotransposons. *Annu. Rev. Genet.* **33**, 479–532 (1999).
36. Hollister, J. D. & Gaut, B. S. Epigenetic silencing of transposable elements: A trade-off between reduced transposition and deleterious effects on neighboring gene expression. *Genome Res.* **19**, 1419–1428 (2009).
37. Ahmed, I., Sarazin, A., Bowler, C., Colot, V. & Quesneville, H. Genome-wide evidence for local DNA methylation spreading from small RNA-targeted sequences in *Arabidopsis*. *Nucleic Acids Res.* **39**, 6919–6931 (2011).
38. Eichten, S. R. *et al.* Spreading of heterochromatin is limited to specific families of maize retrotransposons. *PLoS Genet* **8**, e1003127 (2012).
39. Eichten, S. R. *et al.* Epigenetic and genetic influences on DNA methylation variation in maize populations. *Plant Cell* **25**, 2783–2797 (2013).
40. Tian, E. *et al.* Detection and molecular characterization of two *FAD3* genes controlling linolenic acid content and development of allele-specific markers in yellow mustard (*Sinapis alba*). *PloS One* **9**, e97430 (2014).
41. Thies, W. Schnelle und einfache Analysen der Fettsäurezusammensetzung in einzelnen Raps-Kotyledonen. 1. Gaschromatographische und papierchromatographische Methoden. *Z Pflanzenzüchtg* **65**, 181–202 (1971).
42. Downey, R. K. & Harvey, B. L. Methods of breeding for oil quality in rape. *Can. J. Plant Sci.* **43**, 271–275 (1963).
43. Siebert, P. D. *et al.* An improved PCR method for walking in uncloned genomic DNA. *Nucleic Acids Res.* **23**, 1087–1088 (1995).
44. Ausubel, F. M. *et al.* *Current protocols in molecular biology*. New York: John Wiley. (1995).
45. Livak, K. J. & Schmittgen, T. D. Analysis of relative gene expression data using real-time quantitative PCR and the 2<sup>(-Delta Delta C(T))</sup> Method. *Methods* **25**(4), 402–408 (2001).
46. Clough, S. J. & Bent, A. F. Floral dip: a simplified method for *Agrobacterium*-mediated transformation of *Arabidopsis thaliana*. *Plant J* **16**, 735–743 (1998).
47. Jako, C. *et al.* Seed-specific over-expression of an *Arabidopsis* cDNA encoding a diacylglycerol acyltransferase enhances seed oil content and seed weight. *Plant Physiol.* **126**, 861–874 (2001).
48. Van Ooijen, J. W. JoinMap 4. Software for the calculation of genetic linkage maps in experimental populations Kyazma BV. Netherlands: Wageningen (2006).
49. Van Ooijen, J. W. MapQTL 6: Software for the mapping of quantitative trait loci in experimental populations of diploid species. Kyazma BV: Wageningen, Netherlands (2009).

## Acknowledgements

Breeding and research programs for the condiment yellow mustard are supported by the Growing Forward II-Agri-Innovation Program, Mustard 21 Canada Inc. There is no conflict of interest. We sincerely thank Dr. Rong Zhou and Ms. Dawnne Campbell for seed and yeast fatty acid analysis. The vector pBI101 was kindly provided by Dr. Kevin Rozwadowski at Agriculture and Agri-Food Canada-Saskatoon Research and Development Centre.

## Author Contributions

B.C. conceived the project. F.Z. and B.C. designed research. V.R. performed QTL mapping. F.Z. did cloning and expression studies and analyzed data. F.Z., V.R. and B.C. wrote the article. All authors reviewed the manuscript.

## Additional Information

**Supplementary information** accompanies this paper at doi:10.1038/s41598-017-08317-y

**Competing Interests:** The authors declare that they have no competing interests.

**Publisher's note:** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2017