Evolutionary Pattern of the FAE1 Gene in Brassicaceae and Its Correlation with the Erucic Acid Trait

Xiaojin Sun*, Hui Pang*, Mimi Li, Bin Peng, Haisong Guo, Qinqin Yan, Yueyu Hang

Jiangsu Province Key Laboratory for Plant Ex Situ Conservation, Institute of Botany, Jiangsu Province and Chinese Academy of Sciences, Nanjing, Jiangsu, China

Abstract

The fatty acid elongase 1 (FAE1) gene catalyzes the initial condensation step in the elongation pathway of VLCFA (very long chain fatty acid) biosynthesis and is thus a key gene in erucic acid biosynthesis. Based on a worldwide collection of 62 accessions representing 14 tribes, 31 genera, 51 species, 4 subspecies and 7 varieties, we conducted a phylogenetic reconstruction and correlation analysis between genetic variations in the FAE1 gene and the erucic acid trait, attempting to gain insight into the evolutionary patterns and the correlations between genetic variations in FAE1 and trait variations. The five clear, deeply diverged clades detected in the phylogenetic reconstruction are largely congruent with a previous multiple gene-derived phylogeny. The Ka/Ks ratio (<1) and overall low level of nucleotide diversity in the FAE1 gene suggest that purifying selection is the major evolutionary force acting on this gene. Sequence variations in FAE1 show a strong correlation with the content of erucic acid in seeds, suggesting a causal link between the two. Furthermore, we detected 16 mutations that were fixed between the low and high phenotypes of the FAE1 gene, which constitute candidate active sites in this gene for altering the content of erucic acid in seeds. Our findings begin to shed light on the evolutionary pattern of this important gene and represent the first step in elucidating how the sequence variations impact the production of erucic acid in plants.

Introduction

Erucic acid is one of the major fatty acids present in the oil extracted from members of the family Brassicaceae. Together with other very long-chain fatty acids (VLCFAs; >18C), erucic acid is a precursor of many biologically important compounds, such as waxes [1], sphingolipids [2,3] and triacylglycerols [4]. In Arabidopsis, FAE1 encodes for a b-ketoacyl-CoA synthase (FAE1 KCS) that catalyzes the initial condensation step in the elongation pathway of VLCFA biosynthesis [5].

After the first FAE1 gene was cloned in Arabidopsis via transposon tagging [5], Lassner et al. [6] cloned a cDNA homologous to the Arabidopsis FAE1 gene from jojoba and the role of this cDNA in producing erucic acid was genetically ascertained through genetic transformation of low erucic acid content rapeseed. In addition, two loci, E1 and E2, encoding KCS enzymes were isolated from rapeseed using an oligonucleotide probe based on the Arabidopsis FAE1 gene [7,8]. In Brassica napus, two homologs of the FAE1 gene (Bn-FAE1.1 and Bn-FAE1.2) have been characterized, showing 99.4% nucleotide identity and a two-base deletion resulting in functional loss of the Bn-FAE1.2 gene in the C genome [8]. The LEA trait of rapeseed of the ORO origin is attributed to the substitution of a single amino acid residue, replacing serine with phenylalanine at position 282 of the encoded protein [9,10]. Based on the alignment of amino acid sequences, the putative active site of the KCS, Cys223, His391 and Asn424 were predicted to be crucial for enzyme function [11,12]. These amino acids, together with four conserved histidine residues and six conserved cysteine residues identified by Ghanevati and Jaworski [11], were subsequently found to be present in all of the HEA and LEA rapeseed cultivars analyzed to date [9,10,13]. Thus, the impact of sequence variations on the content of erucic acid in the seeds of Brassicaceae remains unclear. In addition, non-3 integer deletions in a fragment of the FAE1 gene, which are predicted to lead to frameshift mutations and premature termination of translation, were shown to be...
responsible for the low erucic acid trait in rapeseed independent from the point mutation [8,14].

Although FAE1 genes from different Brassicaceae species have been isolated and manipulated extensively through expression in homologous species or in heterologous host systems, the evolutionary history of FAE1 in Brassicaceae remains unclear, and few studies have been focused on oil crops or the model plant species. For example, based on analysis of the evolutionary relationships and sequence similarities among three Brassica species, a closer pedigree relationship was identified between the Brassica A and C genomes than between the A/C genomes and the Arabidopsis genome. Additionally, 18 SNPs have been found in the coding region of FAE1 that may be used as genome-specific markers to differentiate the A and C genomes [15]. Cloning and phylogenetic analyses of FAE1 in Camelina sativa support a history of polyploidization and indicate that C. sativa and C. microcarpa might share a parental genome [16].

Up to 338 genera and 3,709 species are included in Brassicaceae [17]. However, the taxonomy of Brassicaceae has long been controversial because of the often poorly delimited generic boundaries and artificially circumscribed tribes, resulting in a lack of agreement with regard to the number of and boundaries between tribes and genera, giving rise to several systems of classification put forth during the past two centuries [18]. Recently, a comprehensive study was performed to assign 301 genera (94%) of the family to 49 tribes, resulting in a lack of agreement with regard to the number of and boundaries between tribes and genera, giving rise to several systems of classification put forth during the past two centuries [18]. Recently, a comprehensive study was performed to assign 301 genera (94%) of the family to 49 tribes, suggesting an updated framework for the entire family [19]. Based on the combination of molecular work with trichome characteristics, three major lineages have been proposed during the last decade in the core Brassicaceae, which are sister to the basal group (Aethionema) [20]. These lineages are well supported by multiple plastid and nuclear sequences, including phyA [21], ITS [22], nad4 intron 1 [23] and combined molecular datasets (adh, chs, ITS, matK, trnL-F [24] and adh, chs, ITS, matK, nad4 intron 1, ndhF, rbcL, trnL-F [25]).

The content of erucic acid in the seeds of Brassicaceae is genetically variable [26,27]. As the FAE1 gene serves as the key regulator in erucic acid biosynthesis, an evaluation of the diversity and a phylogenetic analysis of the FAE1 gene were performed in this study to elucidate how the plant FAE1 gene originated and evolved. Detection of variations and a correlation analysis between the FAE1 genotype and phenotype were also performed to demonstrate how these genetic variations impact the erucic acid content in plants.

Results

FAE1 Polymorphism

Forty-eight accessions were collected for FAE1 cloning from around the world (the sequences of these accessions are provided with the GenBank accession number of initial letters “KF” in Table 1). Orthologous sequences of FAE1 from Arabidopsis thaliana, Arabidopsis lyrata subsp. kamchatcica, Brassica napus, Brassica rapa, Brassica oleracea, Crambe glabrata, Crambe kralikii, Crambe hispanica subsp. abyssinica, Isatis tinctoria, Lepidium apetalum, Orychophragmus violaceus, Sinapis alba, Sinapis arvensis and Teesdalia nudicaulis were downloaded from GenBank (the GenBank accession numbers are provided in Table 1). In total, 62 accessions, representing 14 tribes, 31 genera, 51 species, 4 subspecies and 7 varieties were used in this study. Among these accessions, the entire coding region of FAE1 could be assessed in 26 accessions, whereas only partial fragments could be obtained in the other 36 accessions due to failure of PCR amplification with all four published sets of primers [14,28-30]. All 96 obtained sequences share high homology with the FAE1 gene of Arabidopsis thaliana (GenBank accession: U29142.1), showing an identity higher than 80%. Among the 11 accessions randomly selected for a heterozygosity analysis of the FAE1 gene, 7 accessions harbor different FAE1 copies numbers, ranging from one copy in Brassica tournefortii, Eruca vesicaria subsp. sativa, Coronopus didymus and Thlaspi arvense to three in Crambe hispanica. This observation suggested that the gene may have undergone copy number evolution. The length of the alignment matrix is 892 bp, corresponding to 431-1,322 bp in the FAE1 protein-coding region beginning with the start codon of Arabidopsis thaliana (GenBank accession: U29142.1), which contains no introns. A total of 489 variable sites and 375 informative sites were found, accounting for 54.82% and 42.04% of the matrix, respectively.

The nucleotide diversity (Pi) increased to 0.1024 and 0.0994 when Aethionema grandiflorum was excluded. The pairwise distances of the FAE1 sequences among ingroup taxa ranged from 0, between 97 pairs of species in Brassicaceae and one species pair each in Rorippa and Cardari, to 32.3% between Aethionema grandiflorum and Lobularia maritima var. benthami. The nonsynonymous to synonymous substitution ratio (Ka/Ks) was determined to be 0.0947.

According to the tribe assignment of Al-Shehbaz et al. [19], the 62 accessions belong to 14 tribes, with the remaining species Lunaria annua being unplaced. In the FAE1 dataset, 6 tribes showed unique character states that differentiated them from the other tribes of Brassicaceae (Table S1). For example, tribes Brassiceae and Lepidieae display 16 and 2 diagnostic sites, respectively (e.g., positions 119: A and 506: C for Aethionema grandiflorum was excluded). The pairwise distances of the FAE1 sequences among ingroup taxa ranged from 0, between 97 pairs of species in Brassicaceae and one species pair each in Rorippa and Cardari, to 32.3% between Aethionema grandiflorum and Lobularia maritima var. benthami. The nonsynonymous to synonymous substitution ratio (Ka/Ks) was determined to be 0.0947.

Evolution of FAE1 and Correlation with Phenotype

The phylogenetic tree of FAE1 was obtained based on the maximum likelihood method (Figure 1; Figure S1). The parsimony and Bayesian inference trees displayed similar topologies to the maximum likelihood tree, which revealed a distinct grouping among the 62 accessions, with Tropaeolum majus and Limnanthes douglasii as outgroup. Five major
Table 1. Collection of Brassicaceae species for evolutionary analysis of the FAE1 gene.

| Tribe              | Genus        | Species                                      | Depositary (Accession); Geographical Origin | Genebank No.* |
|--------------------|--------------|----------------------------------------------|---------------------------------------------|---------------|
| Lepidieae          | Lepidium     | L. campestre (L.) R. Brown                   | GRIN (PI633246); Mongolia                    | FJ907545.1    |
| Coronopus          | C. didymus   | C. didymus (L.) Smith                        | Nanjing Botanic Mem. Sun Yat-Sen. (76A); China | KF664174      |
| Cardaria           | C. daub (L.) Desvaux |                                            | Xinjiang Agricultural University (41A); China | KFO30548      |
| Cardamineae        | Rorippa      | R. indica (L.) Hiern                         | Nanjing Botanic Mem. Sun Yat-Sen. (18A); China | KFO30537      |
| Brassiceae         | Brassica     | B. oleracea var. botrytis L.                 | GRIN (PI115881); India                      | KFO30504      |
|                    |              | B. oleracea var. italic Plenck               | GRIN (PI231210); Italy                      | KFO30505      |
|                    |              | B. oleracea var. gemmifera (de Candolle) Zenker | GRIN (PI209942); Australia                    | KFO30506      |
|                    |              | B. oleracea var. argylodes L                 | GRIN (PI121908); Iran                       | KFO30507      |
|                    |              | B. oleracea var. albiflora kuntze            | GRIN (PI249556); Thailand                    | KFO30508      |
|                    |              | B. napus L.                                  |                                             |               |
|                    |              | B. napus var. napobrassica (L.) Reichenbach  | GRIN (PI263056); Russian Federation          | KFO30512      |
|                    |              | B. elongata Ehrhart                          | GRIN (Ames21299); Iran                      | KFO64168, KFO64169 |
|                    |              | B. carinata A. Braun                         | PGRC (CN40216); Canada                      | KFO664166, KFO664167 |
|                    |              | B. tournefortii Gouan                        | PGRC (CN43445); India                       | KFO664165     |
|                    | Sinapis      | S. alba L.                                   |                                             | AY888040.1    |
clades were found: Clade V is sister to the remainder of the family, and the other clades (I-IV) are nearly equally distantly related. Bootstrap support for these clades was strong in some instances (e.g., greater than 55% for I, II, IV and V) but considerably weaker for Clade III (below 50%). Most of accessions in Clade I and all accessions in Clade II fall into lineage II and lineage I respectively, so these two clades hereafter could be referred to as lineage II and lineage I respectively too. Within Clade I, 28 species of Brassicaceae formed a strongly supported monophyletic group; other members of Clade I include two species each of Sisymbrieae and Isatideae, Orchophragmus violaceus and Goldbachia laevigata. The two species of Thlaspiideae were separated in the FAE1 gene tree: Thlaspi arvense and Thlaspi perfoliatum were distributed in Clades I and III, respectively. Clade II can be divided into two distinct sub-clades. Five species of Cardamineae formed a strongly supported sub-clade, and the species in the other sub-clade represent three tribes, Camelineae, Erysimeae and Lepidiceae, together with one species each of Descurainiaeae and Cardamineae. Thlaspi perfoliatum and Lunaria annua formed a monophyletic Clade (III), though this placement showed moderate-weak support. Clade IV was well supported as monophyletic, including Cochlearia officinalis, Teesdalia nudicaulis and Lobularia maritima. Clade V was comprised of Aethionema grandiflorum and was strongly supported as a sister clade to the remainder of the Brassicaceae family.

The divergence levels (Dxy) between the five groups were calculated, revealing more than 13% of genetic divergence among them (Table 2). Group V was found to be considerably divergent from the other groups, and the nucleotide diversity (π) increased to 0.0679, 0.1063, 0.1920 and 0.2134 between group V and groups I, II, III and IV, respectively, with the diversity between alleles within groups V and IV showing extreme dissimilarity. Similarly, the diversity within a group decreased from 0.1715 in IV, 0.1110 in III and 0.0954 in II to 0.0626 in group I. Slightly different rates of diversity were obtained between the groups and closely related species within
Concentrated Maximum likelihood phylogeny tree (-ln likelihood=30921.31) of Brassicaceae FAE1 showing tribes. Numbers at branches are bootstrap percentages (the values less than 50% are not shown) of ML and MP and posterior probability of BI. Numbers or letters in brackets next to species represent different sequences downloaded from Genbank or different isoforms of the same species. See full Maximum likelihood phylogeny tree in Figure S1.

doi: 10.1371/journal.pone.0083535.g001

the same order (Brassicales). The lowest nucleotide diversity (0.0811) was found between group I and the out-group; group II differed from the out-group with an intermediate level of diversity (0.1383). The diversity values obtained between groups III, IV and V and the out-group were almost the same, at 0.3197, 0.3340 and 0.3671, respectively. The percentage of PS within a group reflects the richness of genetic variation in each group, and the obtained PS% values were 11.10%, 24.33%, 32.51% and 39.01% for groups III, IV, II and I, respectively.

It is notable that the well-known allopolyploids in the genus Brassica, such as Brassica napus, produced two isoforms of the FAE1 gene, with each clustering separately with its parental diploid species Brassica oleracea and Brassica rapa, corresponding to the C-genome of Brassica oleracea and the A-genome of Brassica rapa. The same situation was found in Brassica carinata, with the two isoforms corresponding to the C-genome of Brassica oleracea and the B-genome of Brassica nigra. Conversely, Brassica juncea, another allopolyploid, exhibited only one isoform clustered in the A-genome subclade because another isoform has not been identified and is therefore not included in this analysis. The divergence levels between the A-, B- and C-genomes showed more than 2% of genetic divergence between them, and 11, 45 and 47 mutations were detected as being fixed between the A- and C-genomes, B- and C-genomes or A- and B-genomes, respectively. The separate clustering of isoforms observed for each of Crambe hispanica, Crambe hispanica subsp abyssinica, Crambe filiformis, Raphanus raphanistrum, Brassica elongata, Cardamine parviflora and Capsella bursa-pastoris indicated that the FAE1 genes of these species might have undergone a not particularly distant gene duplication or triplication event, perhaps along with or just after the species differentiation, similar to what has been observed for Camelina sativa [16].

Relationship between the phenotypes and genotypes of FAE1

Our previous work determined the erucic acid contents in the seeds of 60 accessions [27], which ranged from 0 to 55.82% (Table 3). Gaps in the distribution of the erucic acid content, between 9.21% and 10.5%; 18.16% and 20.5%; 28.32% and 31.11%; and 39.8% and 41.1%, allowed us to group the alleles into five operational subclasses of erucic acid content in seeds, defined as low (L), intermediate (M1~M3), and high (H). This categorization of the accessions was strongly supported by an ANOVA analysis (P << 0.001 in a t-test for all pairs between the five categories).

The alleles from the low, intermediate and high erucic acid accessions were not scattered throughout the FAE1 gene tree but grouped together (Figure 1). Therefore, we tested for a significant association between FAE1 sequence variation and erucic acid content variation by analyzing the differentiation (an Fst estimator based on nucleotide diversities [31]) between different phenotypes. The overall differentiation between the phenotypes was highly significant (L, M1, M2 and M3, H; Fst=0.2335, P=0.0062; Table 3), whereas no significant differentiation was detected between any of the pairs of

Figure 1. Concentrated Maximum likelihood phylogeny tree (-ln likelihood=30921.31) of Brassicaceae FAE1 showing tribes. Numbers at branches are bootstrap percentages (the values less than 50% are not shown) of ML and MP and posterior probability of BI. Numbers or letters in brackets next to species represent different sequences downloaded from Genbank or different isoforms of the same species. See full Maximum likelihood phylogeny tree in Figure S1.

doi: 10.1371/journal.pone.0083535.g001
Phenotypic variation in the FAE1 gene is correlated with the content of erucic acid in the seeds, which is suggestive of causal links between the genotype and phenotype.

To further explore the relationship between the observed genetic variation and phenotypes, the seven most conserved motifs within the FAE1 coding sequence were identified using MEME (Figure 2). These motifs encompassed the entire FAE1 encoding region, with no overlap being detected. Fixed mutations in one group reflect a conservative genetic state in that group that is unique from the other groups. Most of the accessions with high or low erucic acid contents were distributed in Clades I and II, and we detected 16 fixed mutations between these two groups (with >70% allele frequency found in one group, but <30% in the other group), revealing the evolutionary history originating from the base (Aethionema grandiflorum) of Brassicaceae (Figure 2). In general, there were two indicated evolutionary pathways for a single mutation to have become fixed between the groups: one is to be diverged in Clades I and II from the existing alleles in the founder accessions (e.g., at positions 37, 38, 70, 111, 112, 142, 152, 182, 213 and 225); and the other is to newly appear in Clades I or II and differentiate itself from the other groups (e.g., at positions 69, 117, 150, 172, 193 and 264). Functional studies, such as through site-directed mutagenesis, could help to determine the active site(s) responsible for the function of FAE1 enzyme.

Discussion

Brassicaceae phylogenetics inferred from FAE1

A comprehensive phylogeny of Brassicaceae based on the multiple plastid and nuclear genes revealed Aethionema as a sister group to the rest of the family (core Brassicaceae), which formed three well-defined major lineages (Lineages I-III), each consisting of several tribes [20-24,32]. The phylogenetic tree of the FAE1 gene presented herein is largely congruent with a previous multiple gene-derived phylogeny, as follows. Clade I includes both tribes of Lineage II (Figure 1; with a well-support bootstrap value of 95) plus Thlaspi arvense (Thlaspiideae) and Goldbachia laevigata (Calepinaceae), which have been considered to belong to in "expanded Lineage II" [31]; Clade II contains the tribes of Lineage I (Figure 1; tribes Cameliinae, Chorisoporeae, Descurainaea, Erysimeae, Cardamine, Lepidaceae). Thlaspi perforiatum (Thlaspiideae) in Clade III and Cochlearia officinalis (Cochlearieae), Teesdalia nudicaulis (Iberideae) and Lobularia maritima var. benthamii (Alyssaeae) are placed in "expanded Lineage III" [31]. Lunaria annua in Clade III is not currently recognized in any of the three Lineages. The basal position of Aethionema is well supported by FAE1 in this study and in previous reports [20-24,32].

In addition to the congruence of the five major clades with the well-recognized major lineages and basal groups, we report some further interesting results. For instance, Descurainia has been previously placed in the tribe Sisymbrieae [33]. Al-Shebaz et al. [17] reallocated this genus to a newly established tribe, Descurainieae, whereas Descurainia was later grouped with genera such as Smelowska on the basis of multiple markers [21,24]. In this study, the separation of Descurainia from the Sisymbrieae tribe and the grouping of Descurainia within the clade of tribe Cameliinae demonstrate the feasibility and rationality of the establishment of the new tribe Descurainieae and suggest a closer relationship of Descurainia with Cameliinae. Additionally, the species within Thlaspiideae were not grouped into the same clade but were dispersed in the phylogenetic trees obtained for the PHYA gene [21] and the FAE1 gene in this study. These results suggest obvious differentiation in Thlaspiideae, engendering the enthusiastic reconsideration of the previous taxonomy of Thlaspiideae, which was mainly based on morphological characters.

Selection on FAE1

The level of interspecific sequence diversity of the FAE1 gene is clearly greater than that of chloroplast genes, though it is comparable to the levels found in some functional nuclear genes of Brassicaceae. The overall nucleotide diversity value of the nuclear-encoded chalcone synthase gene (CHS) reaches 0.1514% or 0.1373% with of without Aethionema grandiflorum, respectively [24]. Beilstein et al. [20,21] examined the levels of diversity in a chloroplast gene and a nuclear gene within a large sample of 113 species of Brassicaceae. The greatest sequence diversity detected in the chloroplast gene.

Table 2. Nucleotide diversity and divergence within and between groups (or families) and polymorphic sites at FAE1.

| Group | Nucleotide diversity | divergence within or between groups | outgroup |
|-------|----------------------|-------------------------------------|----------|
| I     | 0.0626               |                                     |          |
| II    | 0.0927*** 0.1336     | 0.0954                              | 0.0811|0.3963 |
| III   | 0.0667|0.1348               | 0.0996|0.1233 | 0.1110  | 0.1383|0.3805 |
| IV    | 0.0721** 0.1759      | 0.1120** 0.1731                    | 0.1649|0.1706 | 0.1715  | 0.3140|0.3845 |
| V     | 0.0679|0.2466               | 0.1063|0.2296 | 0.1920|0.2323 | 0.2134|0.2546 | 0.3671|0.3926 |
| Polymorphic sites (%)^* | 348(39.01%) | 290(32.51%) | 99(11.10%) | 217(24.33%) | - |

*Indicate a significant difference at P < 0.01 between two groups; the average nucleotide diversity between groups is calculated from all possible pairs between these groups; Limnithes douglasii and Tropaeolum majus are used as out-group species.

^Indicate a significant difference (P < 0.001) between the numbers in different groups by the Chi-square test.

doi: 10.1371/journal.pone.0083535.t002
### Table 3. Erucic acid content in seeds of 58 accessions in relation to genotypes of FAE1.

| Accession               | Erucic acid content 1 | Category** | Fst between categories |
|-------------------------|-----------------------|------------|------------------------|
|                         |                       |            | M1        | M2        | M3        | H        |
| Cardamine hirsuta       | 0                     | L          | 0.0410   | 0.0940   | 0.2007   | 0.3977   |
| Cardamine parviflora    | 0                     | L          |          |          |          |          |
| Nasturtium officinale   | 0                     | L          |          |          |          |          |
| Lobularia maritima var. | 0.24                  | L          |          |          |          |          |
| benthamii               |                       |            |          |          |          |          |
| Orychophragmus violaceus| 0.37                  | L          |          |          |          |          |
| Aethionema grandiflorum | 0.48                  | L          |          |          |          |          |
| Capsella bursa-pastoris | 0.61                  | L          |          |          |          |          |
| Coronopus didymus       | 1.30                  | L          |          |          |          |          |
| Arabidopsis thaliana    | 1.88                  | L          |          |          |          |          |
| Lepidium apetalum       | 2.41                  | L          |          |          |          |          |
| Camelina microcarpa     | 2.53                  | L          |          |          |          |          |
| Camelina sativa         | 3.03                  | L          |          |          |          |          |
| Goldbachia laevigata    | 8.27                  | L          |          |          |          |          |
| Arabidopsis lyrata subsp. kamchatica | 9.21   | L          |          |          |          |          |
| Neslia paniculata       | 10.50                 | M1         | 0.0379   | 0.3255   | 0.3503   |          |
| Erysimum siliculosum    | 11.01                 | M1         |          |          |          |          |
| Cardaria draba          | 11.04                 | M1         |          |          |          |          |
| Cardaria draba subsp. chalepensis | 12.12 | M1         |          |          |          |          |
| Cochlearia officinalis  | 13.77                 | M1         |          |          |          |          |
| Silymbrium loselii      | 16.89                 | M1         |          |          |          |          |
| Silymbrium officinale   | 17.00                 | M1         |          |          |          |          |
| Descurainia sophia      | 18.16                 | M1         |          |          |          |          |
| Erysimum cheiranthoides | 20.50                 | M2         | 0.2351   | 0.2598   |          |          |
| Erysimum silsymbioides  | 20.87                 | M2         |          |          |          |          |
| Isatis tinctoria        | 21.02                 | M2         |          |          |          |          |
| Diplotaxis tenusililqua | 21.33                 | M2         |          |          |          |          |
| Erucastrum gallicum     | 21.91                 | M2         |          |          |          |          |
| Rorippa indica          | 22.32                 | M2         |          |          |          |          |
| Lepidium campestre      | 22.65                 | M2         |          |          |          |          |
| Diplotaxis muralii      | 23.73                 | M2         |          |          |          |          |
| Rorippa dubia           | 26.16                 | M2         |          |          |          |          |
| Myagrum perlphotatum    | 28.32                 | M2         |          |          |          |          |
| Sinapis alba            | 31.11                 | M3         | 0.0216   |          |          |          |
| Erucastrum canariense   | 31.53                 | M3         |          |          |          |          |
| Cheiranthus cheiri      | 31.62                 | M3         |          |          |          |          |
| Sinapis arvensis        | 34.51                 | M3         |          |          |          |          |
| Brassica nigra          | 34.88                 | M3         |          |          |          |          |
| Raphanus sativus        | 35.51                 | M3         |          |          |          |          |
| Thlaspi perlphotatum    | 36.19                 | M3         |          |          |          |          |
| Brassica juncea         | 36.96                 | M3         |          |          |          |          |
| Thlaspi arvense         | 37.79                 | M3         |          |          |          |          |
| Raphanus raphanistrum   | 39.79                 | M3         |          |          |          |          |

| Accession               | Erucic acid content 1 | Category** | Fst between categories |
|-------------------------|-----------------------|------------|------------------------|
|                         |                       |            | M1        | M2        | M3        | H        |
| Brassica napus          | 39.8                  | M3         |          |          |          |          |
| Brassica oleracea       | 41.10                 | H          |          |          |          |          |
| Brassica carinata       | 42.53                 | H          |          |          |          |          |
| Eruca vesicaria subsp. sativa | 43.50 | H          |          |          |          |          |
| Brassica oleracea var. gemmifera | 44.00 | H          |          |          |          |          |
| Lunaria annua           | 44.96                 | H          |          |          |          |          |
| Crabe krátkii           | 45.50                 | H          |          |          |          |          |
| Brassica oleracea var. albiﬂora | 45.81 | H          |          |          |          |          |
| Brassica oleracea var. botrytis | 48.03 | H          |          |          |          |          |
| Brassica tournefortii   | 48.08                 | H          |          |          |          |          |
| Brassica rapa           | 48.32                 | H          |          |          |          |          |
| Brassica oleracea var. italica | 47.94 | H          |          |          |          |          |
| Crabe filiformis        | 51.88                 | H          |          |          |          |          |
| Crabe hispanica subsp. abyssinica | 53.12 | H          |          |          |          |          |
| Crabe hispanica         | 55.82                 | H          |          |          |          |          |
| Overall Fst             | 0.2335                |            |          |          |          |          |

* These data come from our previous work except that of Crabe krátkii and Brassica napus from Kumar and Tsunoda [47] and Velasco et al [48] respectively.

** indicates a significant difference (P<0.01) for the genetic differentiation between different categories.

PLOS ONE | www.plosone.org 7  December 2013 | Volume 8 | Issue 12 | e83535

Evolution of FAE1 and Correlation with Phenotype

rdhF was 3.68% [20], whereas the sequence variation in nuclear phytochrome A (PHYA) was as high as 7.21% [21]. In contrast, a high level of polymorphism and a rapid rate of evolution have been detected in some other nuclear genes, such as plant disease resistance genes. Up to 9.8% nucleotide diversity has been reported for the

| Accession               | Erucic acid content 1 | Category** | Fst between categories |
|-------------------------|-----------------------|------------|------------------------|
|                         |                       |            | M1        | M2        | M3        | H        |
| Brassica napus          | 39.8                  | M3         |          |          |          |          |
| Brassica oleracea       | 41.10                 | H          |          |          |          |          |
| Brassica carinata       | 42.53                 | H          |          |          |          |          |
| Eruca vesicaria subsp. sativa | 43.50 | H          |          |          |          |          |
| Brassica oleracea var. gemmifera | 44.00 | H          |          |          |          |          |
| Lunaria annua           | 44.96                 | H          |          |          |          |          |
| Crabe krátkii           | 45.50                 | H          |          |          |          |          |
| Brassica oleracea var. albiﬂora | 45.81 | H          |          |          |          |          |
| Brassica oleracea var. botrytis | 48.03 | H          |          |          |          |          |
| Brassica tournefortii   | 48.08                 | H          |          |          |          |          |
| Brassica rapa           | 48.32                 | H          |          |          |          |          |
| Brassica oleracea var. italica | 47.94 | H          |          |          |          |          |
| Crabe filiformis        | 51.88                 | H          |          |          |          |          |
| Crabe hispanica subsp. abyssinica | 53.12 | H          |          |          |          |          |
| Crabe hispanica         | 55.82                 | H          |          |          |          |          |
| Overall Fst             | 0.2335                |            |          |          |          |          |

* These data come from our previous work except that of Crabe krátkii and Brassica napus from Kumar and Tsunoda [47] and Velasco et al [48] respectively.

** indicates a significant difference (P << 0.001) for each pairs between 5 categories by t-test.

doi: 10.1371/journal.pone.0083535.0003
Figure 2. Sliding-window plots along the FAE1 gene regions and fixed mutations between Clade I & II. (a) Diversity of the FAE1 gene at silent sites. Windows include 25 silent sites, with successive displacements of 10 sites. The black boxes represent 7 conserved motifs, marked with corresponding ID. (b) 16 fixed mutations between Clade I & II in phylogeny tree of Brassicaceae FAE1. The consensus sequences of each conserved motif are shown with the simplified topology of phylogeny tree of Brassicaceae FAE1 drafted in left. Blue characters in motifs indicate the fixed mutations and the variations of the fixed sites are listed corresponding to each clade in phylogeny tree.

doi: 10.1371/journal.pone.0083535.g002
**Evolution of FAE1 and Correlation with Phenotype**

**Differentiation of FAE1 genotypes and phenotypes**

In *Arabidopsis*, FAE1 encodes for a b-ketoacyl-CoA synthase (FAE1 KCS) that catalyzes the initial condensation step in the erucic acid elongation pathway [5]. Four conserved histidine residues (His302, -387, -391 and -420), six conserved cysteine residues (Cys84, -223, -270, -312, -389 and -460), including the active site at cysteine 223, and an asparagine residue at position 424, required for FAE1 activity, were previously identified by Ghanavati and Jaworski [11,12]. All of these residues were found to be conserved in the 96 Brassicaceae accessions, without substitution and are therefore not responsible for the observed alternations in contents of erucic acid in the different accessions. Although the overall differentiation analysis revealed strong causal links between FAE1 sequence variation and seed erucic acid contents, the correlation for each pair of phenotypes was weak. A possible explanation for this result is that several alternative alleles of FAE1 might confer the same activity, either resulting in the high or low trait. Both the high haplotypes and low haplotypes can be widely divergent; i.e., the FAE1 protein can tolerate a significant number of substitutions while retaining its function, and there are many means of rendering an allele nonfunctional. As additional FAE1 alleles are analyzed, it would be most interesting to identify alleles placed within a clade that lead to a high or low content of erucic acid and determine whether they are functional. One particular aspect of phenotypic variation is worth naming. *Orychophragmus violaceus* encodes an FAE1 protein that is highly similar to that of the other accessions in tribe Brassicaceae (> 87% nucleotide identity), though the contents of erucic acid in the seeds of these accessions are substantially different. The seed erucic acid content of *Orychophragmus violaceus* is close to 0, whereas the Brassicaceae accessions show contents as high as 21%~49%. Such phenotypic variation suggests that the two FAE1 alleles are regulated differently. In support of this hypothesis, heterologous expression of the FAE1 genes of *Orychophragmus violaceus* in yeast resulted in no erucic acid production, indicating that a loss of enzyme activity might be responsible for the low erucic acid trait in *Orychophragmus violaceus* [28].

Deletions in the FAE1 sequence were proven to be responsible for the low erucic acid phenotype, corresponding to secondary mutations independent of point mutation [9,14,36]. A base-pair deletion in the FAE1 sequence is first reported here in *Coronopus didymus*, which would lead to a frameshift mutation and a premature end of the translation. This one-base pair deletion was later shown to be one of the mutations responsible for the low erucic acid trait through heterologous expression in a yeast system [37] and is clearly independent of the previously reported two- and four-base-pair deletions [9,14,36]. Notably, there is another three-base-pair deletion (699-701) found at the base (*Aethionema grandiflorum*) of Brassicaceae. Interestingly, this three-base-pair deletion was also detected in all 32 whole-genome sequenced plants beginning with two members of the Chlorophyta, together with the two outgroup species in the same order, Brassicales. Therefore, this three-base-pair deletion is expected to represent an insertion occurring in the core Brassicaceae, exclusive of the base species. This deletion would introduce a tyrosine residue into the FAE1-encoded protein of all core Brassicaceae species. However, there have been no reports addressing the relationship between this tyrosine insertion and the function of the FAE1 gene before, and based on the high levels of erucic acid observed in the two outgroup species in Brassicales, which do not exhibit tyrosine insertion, no strong evidence is provided of a correlation in the two outgroup species. Thus, it can only be concluded that this tyrosine insertion may serve as an indicator of a trend occurring during evolution from the base to the core of Brassicaceae.

In conclusion, the level of diversity and the available sequence data among different FAE1 alleles provided a rich opportunity to study both evolutionary and genetic aspects of a known functional gene. Indeed, there is great potential for the utilization of evolutionary methods to evaluate diversity, combined with knowledge of gene structure and function to elucidate the molecular mechanisms underlying fatty acid synthesis. By determining the specific amino acid substitutions that alter the composition and content of specific fatty acids, inferences can be made about how the FAE1 gene directs condensation within the elongation pathway of very long-chain fatty acids in plants. These functional studies can, in turn, help to define the reaction mechanism for a membrane-bound condensing enzyme.

**Materials and Methods**

**Plant materials and orthologous sequences**

Forty-eight accessions were collected from around the world, mostly from GRIN (Germplasm Resources Information Network) and PGRC (Plant Gene Resources of Canada) (Table 1). The seeds were germinated at the Institute of Botany, Jiangsu Province and the Chinese Academy of Sciences for DNA extraction. Orthologous sequences of FAE1 from *Arabidopsis thaliana*, *Arabidopsis lyrata* subsp. *kamchatka*, *Brassica napus*, *Brassica rapa*, *Brassica oleracea*, *Crambe glabrata*, *Crambe kralikii*, *Crambe hispanica* subsp. *byssinica*, *Isatis tinctoria*, *Lepidium apetalum*, *Orychophragmus violaceus*, *Sinapis alba*, *Sinapis arvensis* and *Teesdalia nudicaulis* were downloaded from GenBank (the GenBank accession numbers are provided in Table 1). In total, 62 accessions, representing 14 tribes, 31 genera, 51 species, 4 subspecies and 7 varieties were used in this study.

**FAE1 cloning**

For all materials, genomic DNA was extracted from fresh leaves using a modified cetyltrimethylammonium bromide (CTAB) method [38]. Four primer sets were selected from the literatures for *FAE1* gene amplification: TF, GCAATGACGTCCGTTAACGTTAAG, and TR, GGACCGACCCTTTGGAC [29]; FAE1F, AGGATCCTACAATATACCTC, and FAE1R, AGAGAAACATCGTAGCCATCA [28]; F, GCAATGACGTCCATTACGTTAAG, and R, TTAGGACCACGCTTTGGAC [30]; B1F, ATCCGATCATGACGTCGCCGTTAACGTTAAG, and B1R, ATCGAATTCTTAGGACCGACCGTTTTGGAC [14]. As...
to these samples failed to be amplified using the four primer sets, a new primer set (466S, AGATTCAGAGGTGGTACAGG, and 1497AS, TGACATTGCTAGAGCCAC) was designed based on conserved regions with the aid of Oligo primer design software (Molecular Biology Insights, Inc, Cascade, Colorado, USA) using the FAE1 sequence of Brassicaceae deposited in the GenBank database.

Polymerase chain reaction (PCR) amplification was performed using the following program: denaturation at 94°C for 3 min, followed by 35 cycles of denaturation at 94°C for 45 s, annealing at 53–58°C for 30 s and extension at 72°C for 1.5 min. Each 20-μl reaction mixture contained 30 ng of genomic DNA template, 2.5 mmol/L MgCl₂, 1× Mg-free DNA polymerase buffer, 0.12 mmol/L dNTPs, 0.3 μmol/L each primer and 1 U of Taq DNA polymerase. The obtained PCR products were examined electrophoretically in 0.8-1.2% agarose gels. The purified products were sequenced either directly or after cloning into the pMD19-T vector if direct sequencing failed or for each collection. To exclude artificial singleton, the number of clones per individual was added until the haplotype was identified. Bidirectional sequencing was completed by the Beijing Genomics Institute (BGI) using the same primers employed for amplification.

Sequence alignment and data analysis
The sequences were aligned and adjusted manually using Sequencer v.4.5 software (GeneCodes, Ann Arbor, MI, USA), and the nucleotide sequence data for the FAE1 gene were deposited in the GenBank database (Table 1). All sequences, including those of two outgroup species in the same order Brassicales, Tropaeolum majus L. (GenBank accession: AY082610) and Limnanthes douglasii R. Br. (GenBank accession: AF247134), were analyzed using MEGA (5.0 Version [41] and MrBayes (3.2.1 Version [42]) for phylogenetic reconstruction. The two methods of analysis applied in Mega included maximum likelihood (ML) and maximum parsimony (MP), with the bootstrap values being calculated from 1,000 replicates. The corrected Akaike Information Criterion in jModeltest [43] was used to identify the model of evolution that best fit the data for subsequent Bayesian inference analysis. We employed the DnaSp 4.0 program [44] to calculate population-genetic parameters, including the nucleotide diversity (π [45]), the average number of nucleotide differences per site between haplotypes (Dxy [45]), and the Ka/Ks ratio, to complete a sliding-window analysis, to determine silent and replacement sites. Without special explanation, indels were excluded in most calculations; otherwise, total number of polymorphic sites was increased by one site for each indel when length polymorphism sites were considered in these algorithms.

The conserved motifs among the amino acid sequences of the FAE1 region were identified and analyzed using MEME 4.6.1 (Multiple EM for Motif Elicitation [46]). Based on the expectation maximization, the conserved motifs were identified within a group of sequences without a priori assumptions about the alignments. The individual profiles for each conserved motif were assessed, and after tiling, only those conserved motifs showing P-values ≤ 10-4 and no overlap with each other were reported.

Supporting Information

Figure S1. Full Maximum likelihood phylogeny tree (-ln likelihood=30921.31) of Brassicaceae FAE1.

Table S1. Diagnostic sites of tribe, genus and species.

Acknowledgements
We are grateful to Germplasm Resources Information Network, Plant Gene Resources of Canada, germplasm bank of wildspecies, SW China, Prof. Justin Borevitz in the University of Chicago, and Prof. Dunyan Tan in Xinjiang Agricultural University for kindly providing plant materials. We thank the Editor and two anonymous reviewers of Plos ONE for their critical comments on the manuscript.

Author Contributions
Conceived and designed the experiments: XQS YYH. Performed the experiments: XQS HP. Analyzed the data: XQS YYH. Contributed reagents/materials/analysis tools: MML QQY BP QQY HSG. Wrote the manuscript: XQS YYH.

References
1. Post-Beltenmiller D (1996) Biochemistry and molecular biology of wax production in plants. Annu Rev Plant Physiol Plant Mol Biol 47: 405-430. doi:10.1146/annurev.arplant.47.1.405. PubMed: 15012295.
2. Lynch DV, Fairfield SR (1993) Sphingolipid Long-Chain Base Synthesis in Plants (Characterization of Serial Palmitoyltransferase Activity in Squash Fruit Microsomes). Plant Physiol 103(4): 1421-1429. PubMed: 12232036.
3. Merrill AH Jr, Schmelz EM, Dillehay DL, Spiegel S, Shayman JA et al. (1997) Sphingolipids—the enigmatic lipid class: biochemistry, physiology, and pathophysiology. Toxicol Appl Pharmacol 142(1): 208-225. doi:10.1006/tapp.1996.8029. PubMed: 9007051.
4. Downey RK, Röbbelen G (1989) Brassica species. In: G RöbbelenRK DowneyA Ashri. Oil crops of the world. McGraw-Hill, New York. pp 339-362.
5. James DW Jr, Lim E, Keller J, Plooy I, Ralston E et al. (1995) Directed tagging of the Arabidopsis fatty acid elongation1 (FAE1) gene with maize transposon activator. Plant Cell 7: 309-319. doi: 10.2307/3869853. PubMed: 7734065.
6. Lassner MW, Lardizabal K, Metz JG (1996) A jojoba beta-Ketoacyl-CoA synthase cDNA complements the canola fatty acid elongation mutation in transgenic plants. Plant Cell 8(2): 281-292. doi:10.1105/tpc. 8.2.281. PubMed: 8742713.
Arabidopsis family ties: molecular phylogeny and age estimates in the Brassicaceae

doi:10.1007/s00122-007-0685-z. PubMed: 18075728.

Cabbage family affairs: the evolutionary history of Brassicaceae. Trends Plant Sci 16(2): 108-116. doi: 10.1016/j.tplants.2010.11.005. PubMed: 21177137.

Supernetwork identifies multiple events of plastid trnF(GAA) and ndhF sequence data. Mol Biol Evol 28(10): 2731-2739. doi:10.1093/molbev/msr121. PubMed: 21632335.

Evolution of FAE1 and Correlation with Phenotype

genera of Brassicaceae. Chemistry and Industry of Forest Products 31(6): 46-54. (in Chinese with English abstract)

Two genes homologous to Arabidopsis FAE1 co-segregate with the two loci governing erucic acid content in Brassica napus. Theor Appl Genet 96: 852-858. doi:10.1007/s0012200505812.

Functional characterization of beta-ketoacyl-CoA synthase gene from Brassica napus. Plant Mol Biol 46(2): 229-239. doi:10.1023/A:1006651219800. PubMed: 11442206.

Restoring enzyme activity in nonfunctional low erucic acid Brassica napus fatty acid elongase 1 by a single amino acid substitution. Eur J Biochem 269(2): 5625-5631. doi: 10.1111/j.1432-1033.2002.02309.x. PubMed: 12135493.

Mutations in the fatty acid elongase 1 gene are associated with a loss of beta-ketoacyl-CoA synthase activity in low erucic acid rapeseed. FEBS Lett 492(1-2): 107-111. doi:10.1016/S0014-5793(02)02243-8. PubMed: 11248246.

From CoA to saturated fatty acyl-CoA: the role of the fatty acid elongase 1 gene. Theor Appl Genet 116(4): 491-499. doi:10.1007/s00122-007-0685-z. PubMed: 18075728.

Two loci governing erucic acid content in high erucic acid wild species and functional characterization of FAE1 gene from Crambe abyssinica cv. Proplant. Plant Biotechnol J 5(5): 636-645. doi:10.1111/j.1467-7652.2007.00268.x. PubMed: 17565584.

Estimation of levels of gene flow from DNA sequence data. Genetics 132(2): 535-548. doi:10.1093/gerona/32.5-6.551. PubMed: 17325564.

Supplemental information files are available. (in Chinese with English abstract)

Fatty acid composition of the seed oil in a germplasm collection of 94 species in 58 cruciferous species. Genetics 151(2): 911-924. doi:10.1093/genetics/151.2.911. PubMed: 10418169.

Functional characterization of beta-ketoacyl-CoA synthase from Brassica napus. Mol Biol Cell 11(6): 2299-2309. doi:10.1091/0221-1347. PubMed: 10891396.

Theor Appl Genet 116: 491-499. doi:10.1007/s00122-007-0685-z. PubMed: 18075728.

S1388-1981(00)00168-2. PubMed: 11341960.

Functional characterization of beta-ketoacyl-CoA synthase genes from Brassica napus. Plant Mol Biol 46(2): 229-239. doi:10.1023/A:1006651219800. PubMed: 11442206.

Biota selection for the fatty acid elongase 1 gene. Theor Appl Genet 118(10): 1329-1339. doi:10.1007/s00122-007-0685-z. PubMed: 18075728.

Restoring enzyme activity in nonfunctional low erucic acid Brassica napus fatty acid elongase 1 by a single amino acid substitution. Eur J Biochem 269(2): 5625-5631. doi: 10.1111/j.1432-1033.2002.02309.x. PubMed: 12135493.

Arabidopsis thaliana plants and shows a preference for elongating oleic acid to eicosanoid acid. Plant Sci 173(2): 198-205. doi: 10.1016/j.plantsci.2007.05.011.

Phylogenetic analysis of the seed oil in a germplasm collection of 94 species in 58 cruciferous species. Genet Resour Crop Evol 49: 551-566. doi: 10.1023/A:1021210800414.
Evolution of FAE1 and Correlation with Phenotype

Brassica. Genet Resour Crop Evol 45: 371-382. doi:10.1023/A:1008628624867.