Progress toward the identification and stacking of crucial domestication traits in pennycress

Authors

Ratan Chopra[1], Evan B. Johnson[1†], Ryan Emenecker[1†], Edgar B. Cahoon[2], Joe Lyons[3], Daniel J. Kliebenstein[4], Erin Daniels[1†], Kevin M. Dorn[1†], Maliheh Esfahanian[5], Nicole Folstad[1], Katherine Frels[6], Michaela McGinn[5], Matthew Ott[6], Cynthia Gallaher[7], Kayla Altendorf[6], Alexandra Berroyer[5], Baraem Ismail[7], James A. Anderson[6], Donald L. Wyse[6], Tim Umaslov[3], John C. Sedbrook[5], and M. David Marks[1*]

Affiliations

[1] Department of Plant and Microbial Biology, University of Minnesota, St. Paul, MN 55108 USA.

[2] Center for Plant Science Innovation and Department of Biochemistry, University of Nebraska-Lincoln, Lincoln, NE 68588 USA

[3]Covercress, Inc. St. Louis, MO 63132 USA

[4]Department of Plant Sciences, University of California - Davis, Davis, CA 95616 USA

[5]School of Biological Sciences, Illinois State University, Normal, IL 61790 USA

[6] Department of Agronomy and Plant Genetics, University of Minnesota, St. Paul, MN 55108 USA.

[7]Food Science and Nutrition, University of Minnesota, St. Paul, MN 55108 USA

[*]**Correspondence to**: M. David Marks – marks004@umn.edu

[†]**Present addresses:**

E.B.J. Renewable and Sustainable Energy Institute, University of Colorado, Boulder, CO 80309

R.E. Department of Biology, Washington University, St. Louis, MO 63130

E.D. Salk Institute for Biological Studies, La Jolla, CA 92037

K.M.D. Department of Plant Pathology, Kansas State University, Manhattan, KS 66506

Abstract

The oilseed species *Thlaspi arvense* (pennycress) is being domesticated as a new crop that can provide both important ecosystem services and intensify farmland output. Through the use of high throughput sequencing and phenotyping, along with classical mutagenesis key traits needed for pennycress domestication have been identified. Domestication traits identified herein include reduced pod shatter, early maturity, reduced seed glucosinolate levels, and improved oil fatty acid content. By taking advantage of pennycress' close genetic relationship with *Arabidopsis thaliana*, the causative mutations responsible for each of these traits have been identified. These mutations have been used to develop molecular markers to begin to stack the traits into individual lines.

Tremendous gains in crop yields have been achieved through a combination of increased fertilizer use, crop breeding, and improved agronomic practices (2014). However, the long-term sustainability of crop production using current farming practices is threatened by several factors. In intensive cropping systems, such as the corn-soybean rotation in the Midwestern United States, only 40-60% of the applied nitrogen fertilizer is taken up by plants (Wortmann *et al.* 2011). Much of the remaining nitrogen is released as greenhouse gases or pollutes surface and ground waters resulting in the eutrophication of streams and lakes, the creation of dead zones in coastal regions, and the contamination of wells in rural communities (Robertson and Vitousek 2009). Another looming threat to agricultural production is the emergence of herbicide tolerant weeds, including weeds tolerant to glyphosate (Heap 2014). The use of glyphosate resistant crops over the past twenty-five years enabled no-till or reduced-till agriculture that helped to reduce soil erosion (Price *et al.* 2011). Due to the increasing prevalence of herbicide resistant weeds farmers may return to deep tillage to reduce weed pressures, which will hasten soil erosion (Price*, et al.* 2011).

To confound the above issues, the world population is expected to grow to over nine billion by the year 2050 (Gerland *et al.* 2014). To feed this growing population, it has been estimated that food production must increase by an average of 44 million metric tons per year (Tester and Langridge 2010). In addition, global warming due the release of CO_2 from fossil fuels is posing another threat (Anderson *et al.* 2016). To reduce this threat demand is increasing for plant-based renewable feedstocks for the production of biofuels and bioproducts (Ho *et al.* 2014). However, this leads to the concern that using traditional crops for biofuel and bioproducts formation will threaten future food security (Harvey and Pilgrim 2011). One solution is to intensify farmland output (Tilman *et al.* 2011).

The deployment of cover crops during the fallow period between the growth of summer crops can improve water quality and reduce the threat of herbicide tolerant weeds (Dunn *et al.* 2016). However, traditional cover crops

such as winter rye often need to be terminated before maturity to allow the planting of summer crops (Roesch-McNally *et al.* 2017). Thus, such cover crops do not enhance food or feedstock production. The winter annual weed, pennycress (*Thlaspi arvense* L. also known as field pennycress and referred to as "pennycress" hereafter), is being domesticated as a new alternative cover crop (DeHaan *et al.* 2016, Isbell 2009, Jordan *et al.* 2007, Phippen and Phippen 2012, Sedbrook *et al.* 2014). As a winter cover, pennycress can utilize excess nitrates before they escape into the environment and can suppress the growth of spring weeds (Johnson *et al.* 2015, Johnson *et al.* 2017, Weyers *et al.* 201). Importantly, pennycress can be harvested for its oilseeds (Fan *et al.* 2013, Moser 2012, Moser *et al.* 2009a, Moser *et al.* 2009b). Thus, pennycress has the potential to intensify farm output by producing a new crop on land that is temporally held fallow, such as much of the tens of million hectares of land currently undergoing the corn/soybean rotation in the Midwestern United States (Hart 2015).

The selection of pennycress as a target for domestication as a new cover crop was based on positive natural traits such as extreme winter hardiness (-25°C), high seed yields for a wild species (1,000-2,000 kg/ha), and seeds rich in oil (30-35%) and protein (25-27%) (Warwick *et al.* 2002). Pennycress produces an oilseed suitable for making biofuels and bioproducts (Fan*, et al.* 2013, Moser 2012, Moser*, et al.* 2009a). However, the oil and associated seed meal are not suitable for human or animal consumption. The oil is enriched with erucic acid, which has been considered to be unsuitable for human consumption(Bell 1982). In addition, the seeds contain glucosinolates, which are considered to be anti-nutritional for humans and non-ruminant animals (Wittkop *et al.* 2009). Furthermore, the mature seedpods are prone to breakage or shatter resulting in pre-harvest seed loss. Lastly, in many regions, pennycress matures late enough to cause delays in the planting of summer crops. Successful domestication of pennycress requires the elimination of these negative traits.

Herein, we show that the rapid domestication of pennycress can be facilitated by its close relationship to the model plant Arabidopsis (Sedbrook*, et al.*

2014). Like Arabidopsis, pennycress is self-fertile and has a relatively small non-repetitive diploid genome. Genome sequencing and analyses revealed a mostly one-to-one functional correspondence between single *Arabidopsis* genes and single pennycress ortholog genes (Chopra *et al.* 2018b, Dorn *et al.* 2013, Dorn *et al.* 2015). Importantly and directly related to the domestication of pennycress, recessive Arabidopsis mutants have been described that eliminate weedy and other agronomically undesirable traits (Provart *et al.* 2015, Sedbrook*, et al.* 2014). The goal of this study was to identify similar mutants in pennycress with traits considered to be crucial to the domestication of any plant species (Abbo *et al.* 2014). Using classical mutagenesis, we describe the isolation and characterization of pennycress mutations conferring crucial domestication traits such as early flowering/maturity, reduced seedpod shatter, reduced glucosinolates, and reduced erucic acid along with reduced polyunsaturated fatty acids resulting in the creation of canola-like high oleic oil. The causative mutations responsible for each of these traits were identified. These mutant sequences were used to develop molecular markers that have been successfully employed to stack all of the mutations into a single line.

Results

Early flowering/maturing pennycress

For pennycress to fit between the rotations of traditional summer crops it needs to mature in the spring without disrupting the planting of the following summer crop (De Bruin and Pedersen 2008). To identify Arabidopsis-like mutants that flower and mature early, pools of mutagenized population of M_2 seeds were planted in the field space adjacent to the University of Minnesota St Paul campus (Chopra*, et al.* 2018b). During the following spring, plants in the field that flowered early were tagged and followed through to maturity. Sixty-one independent isolates that flowered and matured early compared to the wild-type were identified during the primary screen. Nineteen of the 61 early flowering mutants were confirmed for the inheritance of earliness traits in the following generation. We further focused our efforts on one line, A7-25, that flowered and

145 began producing seedpods 10-14 days earlier than the wild-type (compare **Figs.**
146 **1a**, **b**). In another site-year in St. Paul, this mutant line matured approximately
147 one week ahead of the wild-type allowing an early harvest in the field conditions
148 while demonstrating normal plant stature and yield (**Table. S1**)
149 To test the hypothesis that the A7-25 mutant carried a mutation in a
150 previously characterized in Arabidopsis gene, whole-genome sequencing (WGS)
151 was performed. It was found that A7-25 contains a mutation in the candidate
152 pennycress ortholog of the Arabidopsis *EARLY FLOWERING 6* (*ELF6*) (referred
153 to as *Ta-ELF6-1* in **Fig. 1c**, and **Data S1-3**) (Jeong *et al.* 2009, Noh *et al.* 2004).
154 In Arabidopsis, ELF6 activity de-represses the expression of the floral inhibitor
155 *FLOWERING LOCUS C* (*FLC*) gene by removing methyl groups from lysine 27
156 of histone 3 proteins associated with the *FLC* locus (CrevillÃ©n *et al.* 2014). *FLC*
157 controls flowering time in pennycress(Dorn *et al.* 2017). The reduced
158 demethylase activity in *elf6* mutants results in reduced *FLC* expression, which
159 hastens flowering. As shown in **Fig. 1c** and **Data S3** the mutation in *Ta-ELF6-1*
160 results in the substitution of a tyrosine (Y) residue for a conserved histidine (H)
161 (Noh*, et al.* 2004). This substitution is predicted to disrupt the formation of the
162 iron-binding site in the Jumonji C (JmjC) domain that is required for histone 3
163 demethylase activity in many plant species(Lu *et al.* 2008). To determine if the
164 early flowering phenotype in pennycress co-segregates with the mutation, an F_2
165 population derived from a cross between the mutant and wild-type (Ames23761)
166 was characterized. Statistical analysis of each genotype class in the F_2
167 population showed that plants homozygous for the *Ta-elf6-1* mutation flowered
168 earlier than either homozygous wild-type or heterozygous plants (*p*-values
169 0.0002 and 0.0021, respectively (**Fig. 1d**).
170 All studies with Arabidopsis on the effects of the *elf6* mutation on *FLC*
171 expression have been under controlled growth conditions. To determine if a
172 similar phenomenon occurs in field grown pennycress, we analyzed the effect of
173 the mutation on pennycress *FLC* expression prior to vernalization in the fall. As
174 shown in (**Fig. 1e**), *FLC* expression was reduced by 30.5±0.09% in the early
175 flowering A7-25 line compared to the wild-type grown in the same trial. These

results show that *ELF6* functions the same under field conditions for pennycress as under controlled growth conditions for Arabidopsis.

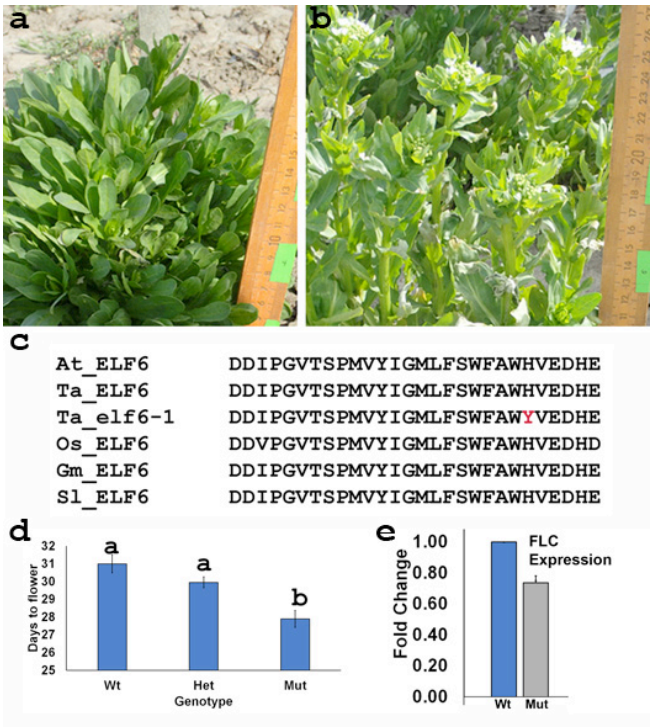


Figure. 1. Early flowering pennycress mutant. Images of field grown wild-type (**a**) and early flowering *Ta-elf6-1* mutant (**b**) plants taken on the same date. (c) Comparison of amino acid sequences within the JmjC domain of *ELF6*-like genes highlighting the residue altered in the pennycress mutant (red). (**d**) Average days to flowering in greenhouse conditions for an F_2 population derived from a cross between wild-type and *Ta-elf6-1* plants. Allele-specific markers were used to assess the *elf6-1* genotype of members in the F_2 population. Note: Letters indicate significant differences based on pair-wise Tukey test. (**e**) qPCR analysis of *FLC* expression in wild-type and early flowering *Ta-elf6-1* mutant plants using RNA from field grown plants collected in the fall of 2017. *FLC* expression values were normalized using a ubiquitin probe (error bar denotes standard deviation). Abbreviations: At – *Arabidopsis thaliana*, Os – *Oryza sativa* (rice), Gm – *Glycine max* (soybean), Sl-*Solanum lycopersicum* (tomato).

Reduced seedpod shatter pennycress

Pennycress seedpods exhibit premature breakage by wind or during mechanical harvest resulting in yield losses of over 50%. To identify pennycress mutants with

196 reduced seedpod shatter, fields of M_2 plants maintained past peak maturity were
197 screened for individuals exhibiting reductions in pod breakage or shatter
198 (compare **Figs. 2a, b).** Five individual recessive mutants have been identified,
199 along with a sixth line identified in a growth room screen, in which the trait
200 successfully passed through multiple generations.
201 Pennycress produces a flat circular seedpod called a silicle that has a
202 visibly distinct morphology compared to the long rod-shaped seedpods known as
203 siliques produced by Arabidopsis and other members of the Brassicaceae plant
204 family (**Fig. 2c)**. Therefore, it was unknown if genes required for pod formation in
205 Arabidopsis would be the same as those controlling pod formation in pennycress.
206 To determine if the pennycress mutants contained mutations in genes known to
207 be associated with the formation of the separation zone in Arabidopsis, which
208 mediates pod breakage, the six lines were subjected to WGS. Three of the lines
209 were found to have mutations in the pennycress candidate ortholog of the
210 Arabidopsis *INDEHISCENT* (*IND*) gene (**Fig. 2d, Data S1-3)** (Liljegren *et al.*
211 2004). The analysis of the remaining three lines will be considered elsewhere.
212 *IND* encodes a *bHLH* transcription factor that regulates the formation of
213 separation layers of cells contained within the value margin, which flanks the
214 septum (also called a replum) that divides the seedpod into to two halves
215 (Liljegren*, et al.* 2004). The *Ta-ind-1* allele contains a mutation outside of a highly
216 conserved bHLH domain, while the *Ta-ind-2* flanks this region and *Ta-ind-3* has a
217 mutation in the conserved bHLH domain (**Fig. 2d and Table S2**). The seedpods
218 of both *Ta-ind-2* and *Ta-ind-3* failed to cleanly break on harvest signifying non-
219 functional dehiscence zones. However, *Ta-ind-1* seedpods did break at the
220 septum to release seeds on harvest, but compared to wild-type, required more
221 force to break open the seedpods (**Fig. 2d**). As expected, *Ta-ind-1* retains a
222 separation layer (**Fig. 2e**) that is not present in *Ta-ind-2* or *Ta-ind-3* (**Fig. 2f** and
223 **Fig. S1**). The finding of three independent pennycress *ind* alleles showing
224 reduced pod breakage strongly supports the causation of these mutations of the
225 observed phenotype. Of the three alleles, *Ta-ind-1* should have the most

agronomic utility, as the shatter-less phenotype of *Ta-ind-2* and *Ta-ind-3* reduces seed harvestability with conventional combines.

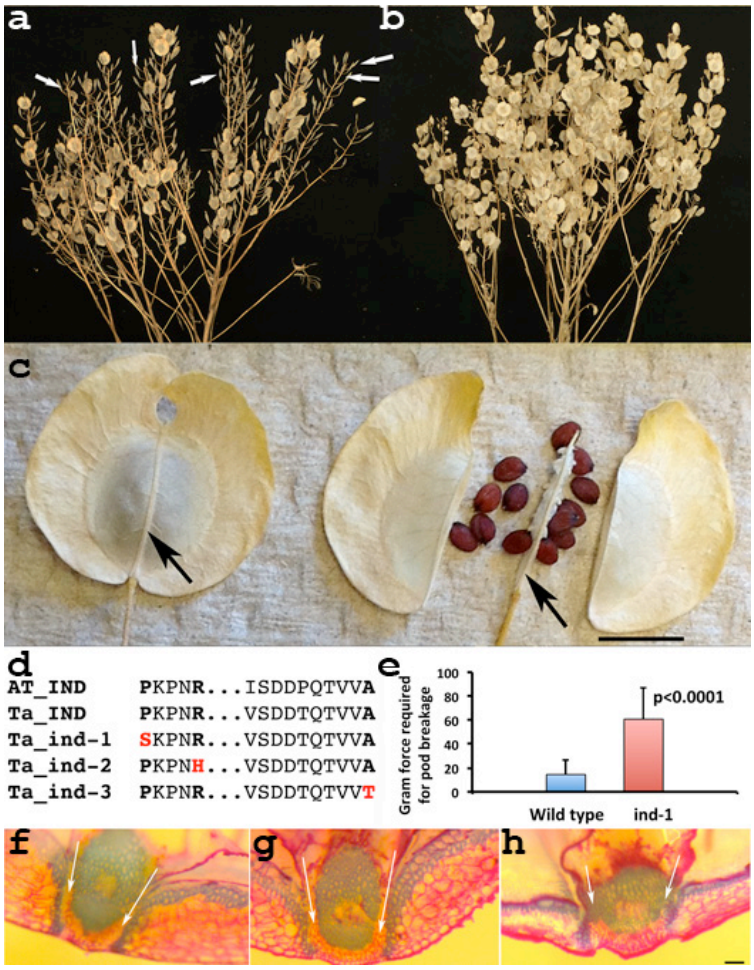


Figure 2. Isolation and characterization of pennycress seed pod mutants. (**a**) Field grown wild-type plant compared to a reduced shatter mutant plant (**b**). (**c**) An intact seedpod compared to a broken seedpod with released seeds (both are wild-type). The arrows show the septum of an intact pod (left) and the septum structure that remains after pod shatter. The arrows highlight the septum remaining after pod shatter. (**d**) IND amino acid sequences derived from Arabidopsis (At-IND) and wild-type pennycress (Ta-IND) compared to three mutant pennycress alleles (*ind-1*, *ind-2* and *ind-3*). The red letters highlight the amino acid substitutions. (**e**) Gram force needed to break open pods from wild-type and *ind-1* plants. Std. dev. error bars are shown. Freehand sections through wild-type (**e**), *ind-1* (**f**), and *ind-2* (**g**) seedpods. The arrows highlight the separation layers in wild-type and *ind-1*. This layer is missing in *ind-2* and *ind-3* (Fig. S2).

241 The darker stained blue regions highlight lignified cell layers. The size bars in **c** and **h**
242 represent 0.5 cm and 50 µm, respectively.
243
244
245 **Low glucosinolate pennycress**
246 Like Arabidopsis and other brassicas, pennycress seeds contain
247 glucosinolates (~100 µmol/g of seeds), almost all in the form of sinigrin (Warwick*, et al.* 2002). Most of the glucosinolates remain in the meal after the extraction of
249 oil from the seeds. This seed meal has added value if it can be used as an
250 animal feed supplement, however high amount of glucosinolates in the meal can
251 negatively affect animals (Wittkop*, et al.* 2009). Thus, a search for pennycress
252 lines with reduced glucosinolates was conducted. As described in a previous
253 report, 15,000 lines were scanned using Near Infra-Red Spectroscopy (NIRS) to
254 estimate seed composition traits in the intact seeds (Chopra *et al.* 2018a). The
255 analysis of these scans identified one line predicted to have very low
256 glucosinolate levels, which was confirmed via wet lab assays for total seed
257 glucosinolates (**Fig. 3a**). Interestingly, numerous individuals indicated that the
258 mutant seeds have a pleasant, nutty flavor that has not previously been reported
259 for seeds from plants in the Brassicaceae plant family. Seeds from the
260 Brassicaceae family typically are characterized to have flavors that range from
261 bitter, mustardy, garlicky, to hot. Hereafter, this pennycress mutant is referred to
262 as “nutty”. The leaves of nutty also lacked the garlicky smell that accounts for
263 pennycress also being called “stinkweed”.
264 In Arabidopsis, the biochemical pathway leading to the formation of
265 various glucosinolates including sinigrin has been well characterized(Sonderby *et*
266 *al.* 2010). Thus, we conducted WGS on nutty and searched for a mutation in a
267 gene known to be involved in glucosinolate biosynthesis in Arabidopsis. This led
268 to the identification of a pennycress mutation in a gene that is equally related to a
269 family of three tandemly linked genes in Arabidopsis called *ALKENYL*
270 *HYDROXALKYL PRODUCING 1, 2, and 3* (*AOP1,2,3*)(Kliebenstein *et al.* 2001).
271 The analysis of an F_2 from a cross between wild type and mutants showed that

the reduced glucosinolate nutty phenotype was tightly linked to the mutation in the AOP-like sequence (**Fig. S2**). HPLC analysis of glucosinolates from nutty confirmed the reduction in sinigrin and also revealed a mild excess accumulation of 3-methylsulfinylpropyl and 3-methylthiopropyl glucosinolates, which are precursors to sinigrin in the biosynthetic pathway, supporting the hypothesis that the pennycress *AOP-like* gene encodes an enzyme with *Arabidopsis AOP2* alkenyl producing activity converting 3-methylsulfinylpropyl glucosinolate to sinigrin (**Fig. 3d**) *(Kliebenstein, et al. 2001).* Thus, the mutant gene was named *Ta-aop2*-1. Of note, all three glucosinolates shown in the HPLC profile shown in **Fig. 3c** can be detected in the glucosinolate profile shown in **Fig. 3a**. If there was a simple blockage in the pathway one would predict that the total levels of glucosinolates would be the same in both mutant and wild type, however, there is a clear reduction in glucosinolates in nutty. Additional work is needed to further understand the molecular mechanisms responsible for the overall reduction in glucosinolates in this mutant.

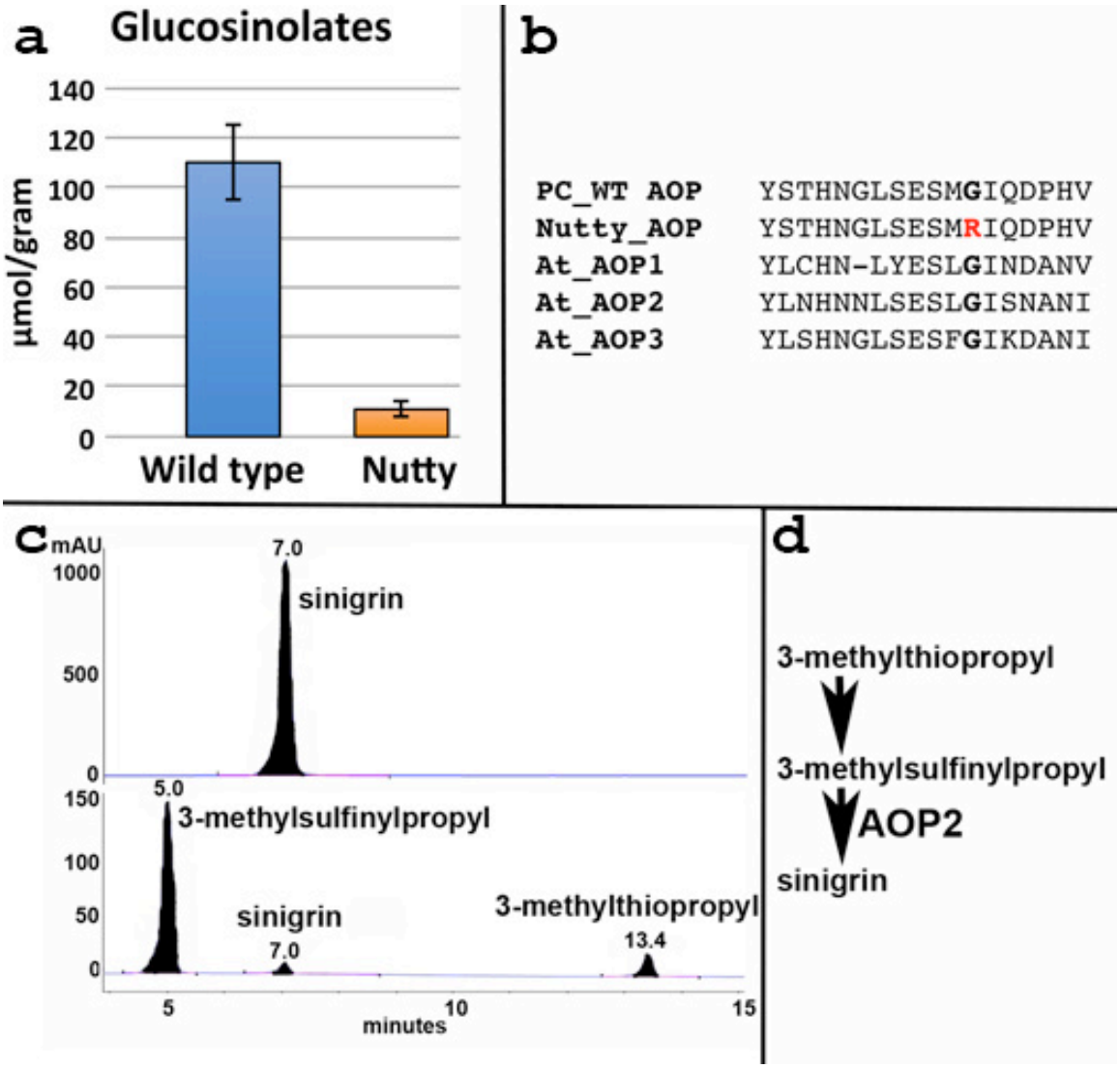


Figure 3. Comparison of glucosinolates in wild-type and Nutty (*aop2-1*) mutant pennycress. (**a**) Spectrometric quantitation of total glucosinolates in wild-type and the nutty mutant. Error bars highlight Std. Dev. (**b**) AOP-like amino acid sequences derived from Arabidopsis (At-AOP 1 through 3) and wild-type pennycress AOP2 (PC_WT AOP) compared to the nutty mutant (Nutty-AOP). (**c**) HPLC analysis of desulfated glucosinolates from wild-type and the nutty mutant, top and bottom, respectively. Peaks are labeled with compound name and the time (minutes) of flow through the HPLC. (**d**) Terminal biochemical steps leading to the synthesis of the glucosinolate sinigrin. The step mediated by the enzyme encoded by *AOP2* in Arabidopsis is shown.

Improving fatty acid composition of pennycress

The fatty acid composition of pennycress oil is similar to that of rapeseed containing over 35% erucic acid (**Fig. 4a**) (Claver *et al.* 2017, Moser*, et al.* 2009a). Oils containing such high levels of erucic acid are not considered fit for human or animal consumption(Knutsen *et al.* 2016). Therefore, a search was made for lines with reduced erucic acid. This search was aided by the knowledge that in Arabidopsis and other species, the biosynthesis of erucic acid is mediated by a single enzyme, *FATTY ACID ELONGATION1* (*FAE1*), which in two steps converts oleic acid with eighteen carbons and one double bond (18:1) to erucic acid with twenty-two carbons and one double bond (22:1) (James *et al.* 1995). Similar to the detection of glucosinolates, the quantity of erucic acid along with other types of fatty acids can be estimated using NIRS(James*, et al.* 1995). Thus, the same 15,000 NIRS scans that were used to identify the reduced glucosinolate line were used to search for lines with reduced erucic acid. This analysis identified two independent lines with greatly reduced erucic acid. These results were verified via gas chromatography (GC), which showed that both lines only contained trace levels of erucic acid (**Fig. 4b**). The candidate pennycress *FAE1* alleles in these reduced erucic acid lines were amplified via PCR and sequenced. Both lines contained independent nonsense mutations leading to premature stops in the predicted pennycress FAE1 protein sequences (**Data S1-3**). These results were further confirmed by WGS of these lines. The finding of two strong loss of function alleles for the candidate pennycress *FAE1* ortholog giving rise to the expected phenotypes strongly supports causation of the mutations.

Pennycress *fae1* mutants contain the desired reduction in erucic acid and a desirable elevation in oleic acid. They also accumulate higher levels of fatty acid with extra double bonds referred to as polyunsaturated fatty acids (PUFAs). In particular, linolenic acid with 3 double bonds (18:3) was elevated. This fatty acid belongs to the omega3 class of fatty acids, which are associated with reduced heart disease, reduced bone fracture risk, and in reduced childhood obesity (Perng *et al.* 2014, Rajaram 2014). However for many applications, PUFAs are not desirable as the extra double bonds reduce the stability of the oil,

thus shortening the shelf life of the products that contain the oil (Gordon 2001). Therefore, a search was made for lines containing reduced levels of PUFAs using the same NIRS screen. We identified two mutants with predicted reductions in both linoleic and linolenic acids. These reductions were verified using GC analysis (**Fig. 4c**). Both lines contained similar levels of erucic acid as the wild-type, but the reductions in PUFA were compensated for by an increase in oleic acid. The fatty acid profiles of these two lines resembled those from Arabidopsis *reduced oleate desaturation1* (*rod1*) mutants(Lu *et al.* 2009). This was confirmed via the PCR amplification and sequencing of the candidate pennycress genes (**Data S1-3**). These results were further confirmed by WGS of these lines.

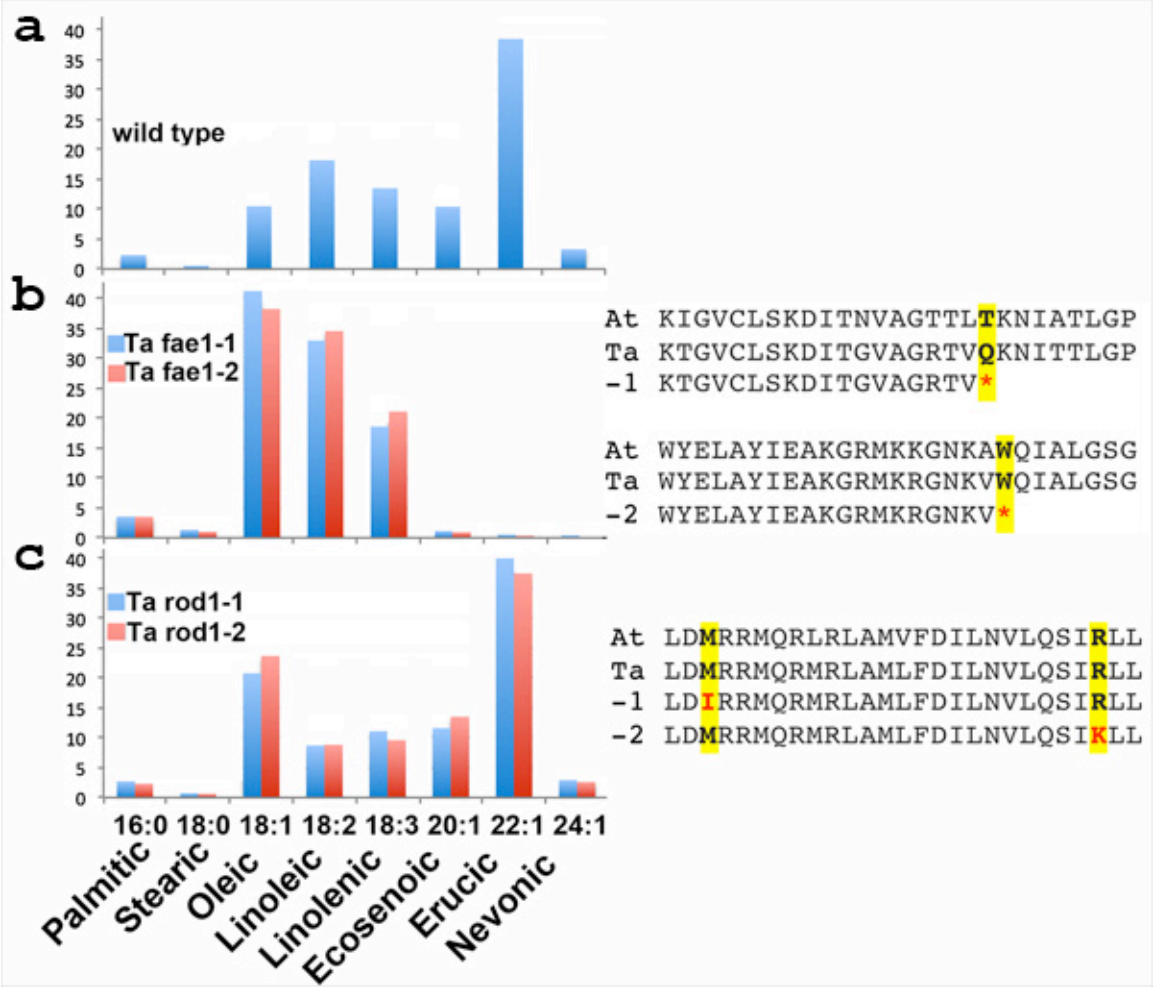


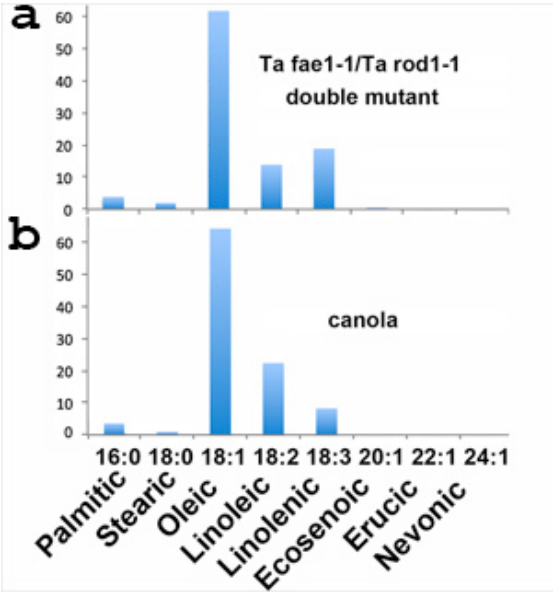
Figure 4. Comparison of wild type fatty acid profiles compared to those of pennycress *fae1* and *rod1* mutant lines. (**a-c**) Profiles of wild-type pennycress, pennycress *fae1* and *rod1* mutants. The amino acid sequences derived from Arabidopsis and wild-type pennycress compared to mutant pennycress alleles are shown to the right. The red letters highlight the amino acid substitutions. Fatty acid names along with carbon chain lengths and the degrees of desaturation are shown across the bottom.

Stacking traits to develop domesticated pennycress

This study has identified pennycress lines that harbor crucial domestication traits including early maturity, reduced shatter, reduced erucic acid and reduced PUFAs. Importantly, none of the described mutations reduced plant stature or predicted yield (**Table S1 and Fig. S3**). Knowledge of the causative mutations in each of the lines has been used to develop molecular probes that can follow each of the traits in segregating populations at the molecular level (**Table S3-4** and **Data S1-3**). To begin to combine these traits, a cross was made between *Ta-fae1-1* and *Ta-rod1-1*. KASP markers were designed for the specific mutations in each of these mutants (see Materials and Methods). Using this approach (**Fig S4**) it was possible to identify the double mutants in an F_2 population at the seedling stage over two months before they began to produce seeds. The fatty acid composition of the oil from these double mutants was very high in oleic acid, low in erucic acid, with reduced levels of in PUFAs. Overall, the fatty acid profile of the double mutant pennycress oil closely resembled the profile of canola oil (**compare Figs. 5a and 5b**) (Dyer *et al.* 2008).

Mutation specific KASP markers have been used to identify additional double and triple mutant combinations in F_1 and F_2 individuals derived from various crosses between the pennycress mutants (**Table S5**). This includes a line that is homozygous for the combined *Ta-fae-1*, *Ta-rod-1*, and *Ta-aop2-1* mutations. This line produces seeds that are essentially equivalent to those from the so-called “double 0” canola that similarly lack erucic acid and contain low levels of glucosinolates(Bell 1982). Additional crosses have generated an F_1

379 plant that is heterozygous for all five of the desirable traits described in this report
380 (**Fig S5**). Progenies from this F_1 plant will be genotyped to identify a line
381 homozygous for all five mutations. This stacked line will represent the first-
382 generation domesticated pennycress variety and will be subjected to field-testing
383 ahead of potential commercialization.
384



385
386
387

388 **Figure 5. Comparison of fatty acid profiles of pennycress fae1/rod1 double**
389 **mutant compared to canola oil.** (**a**) Double mutant, (**b**) canola. Fatty acid
390 names along with carbon chain lengths and the degrees of desaturation are
391 shown across the bottom.
392
393

394 **Discussion**
395
396

397 The traits described in this report should allow domesticated pennycress to
398 intensify farm output by utilizing land from fall to spring that is not currently in
399 production. The early maturing pennycress described in this report will better fit
400 into existing crop rotations. The reductions in pod shatter will greatly improve

harvestable yield, and the reductions in glucosinolates and improvements in oil quality will enhance the value of the oilseeds. The rapid progress in identifying these traits in pennycress was due to reduced gene redundancy, which facilitated the creation of pennycress mutants that closely resemble similar Arabidopsis mutants(Chopra*, et al.* 2018b, Sedbrook*, et al.* 2014).

The causative mutations responsible for the domestication traits described in this study have been well studied in *Arabidopsis*(Provart*, et al.* 2015). Knowing the causative nature of the mutations responsible for the pennycress domestication traits is important for several reasons. The information can be used to stimulate additional basic research on the molecular mechanisms that control these traits. This is especially true for the *Ta-AOP2* gene. The enzyme encoded by this gene catalyzes the last step in the glucosinolate pathway in pennycress. Through some unknown mechanism the mutation in this gene results in reduced metabolic flux through the pathway. The information also can be used to improve current crops or to help guide the domestication and creation of new crops. For example, *ELF6* now becomes good target for hastening maturation of related *B. napus* (canola). Likewise these results show that simultaneously reducing in both *ROD1* and *FAE1* activity can greatly increase oleic acid content above that found for either of the single mutants. This should have wide ranging applications in modifying the seed oil fatty acid content of many plants. Importantly, the causative mutations can be used as molecular markers as in the pipeline shown in **Fig. S4** to combine the pennycress traits into single lines. Using this pipeline, single plant lines harboring all the traits described in this report will be ready for field-testing within the next two years.

The traits described in this report were created using classical mutagenesis. Classical mutagenesis has long been used to induce useful mutations in crop plants (Ahloowalia *et al.* 2004, Oladosu *et al.* 2016). For this work we used EMS as a mutagen (Chopra*, et al.* 2018b). Treatments with EMS generate up to thousands of mutations in each mutagenized line creating a large pool of mutant genes in a relatively small population. This increases the odds of

finding mutants of interest as described in this study. As previously described, these pools mutations can be used to mine for additional desirable traits (Chopra*, et al.* 2018b). On the downside these lines may carry deleterious mutations. However, many of these can be identified from their known affects in Arabidopsis and can be removed by marker-assisted breeding(Chopra*, et al.* 2018b).

In theory, it would have been faster to use gene-editing approaches to create the domestication traits described in this report. In fact, we have created an alternative zero erucic acid pennycress line using CRISPR-Cas9 mediated gene editing (McGinn *et al.* 2018). However, successful utilization of gene editing requires precise knowledge about the genes one wishes to target. This knowledge is not always available. For example, *ELF6, ROD1* and *AOP2-like* were not on the original list of gene targets for improving pennycress(Sedbrook*, et al.* 2014). Furthermore, the creation of partial loss of function mutations, such as that responsible for the desirable *Ta-ind-1* allele, requires prior knowledge before gene editing can be employed. To balance speed and discovery, we will continue to utilize both classical breeding to stack desirable traits and gene editing to produce desirable mutations in existing elite pennycress breeding lines.

There are many obvious improvements that can be made to enhance the potential success of pennycress as a new crop. For example, the dark colored seed coats contain high levels of oxidized tannins, which can interfere with protein uptake in the intestinal tracts of animals that are fed pennycress seed meal(Butler 1992). In a previous report, pennycress mutants lacking tannins have been described, and this trait is currently being added to the stacked line(Chopra*, et al.* 2018b). Additional screens are underway to identify second-generation domestication traits such as reduced seed dormancy, increased yield, increased seed size, increased oil content and improved protein quality. Many of these traits will aid in crop establishment and improve upon current net-economic returns, which increase the likelihood that pennycress will be adopted (Ott *et al.* 2019).

The identification and stacking of domestication traits represent the first step toward the creation a new crop. Concurrently with the ongoing work

described in this report, agronomists and ecologists are working to develop planting recommendations and to understand the effects of growing pennycress on the landscape. In addition, food scientists are beginning to study the potential of pennycress as a new food source for human consumption. For example, the "double zero" - like pennycress seeds lacking erucic acid with greatly reduced glucosinolate will be used for various analyses including feeding trials. The canola-like oil will also provide an important new feedstock for the development biofuels and bioproducts.

Newly domesticated pennycress has the potential to intensify farm output by being grown during the fallow period between existing summer crops. Importantly, pennycress grown during the fallow seasons has been shown to provide much needed ecosystem services including nitrogen uptake and the suppression of spring weeds, which will enhance the sustainability of current agricultural practices. We believe the combined oilseed production and ecosystem rewards will greatly outweigh the remaining challenges in developing pennycress as a new cash cover crop.

Materials and Methods

Screening of mutant pennycress populations for important traits

We previously described the creation of a large pennycress mutant populations using ethyl methane sulfonate (EMS) and other mutagens (Chopra*, et al.* 2018b). These represent the test populations used to search for pennycress mutants similar to Arabidopsis mutants that exhibit agronomically desirable phenotypes.

Seed Source, DNA isolation, Sanger Sequencing, Whole-Genome Sequencing

The source of the pennycress seeds for development of genomic resources was previously described in Dorn et al. (Dorn*, et al.* 2015). Mutants characterized in this report were derived from EMS treated seeds as described in Chopra et al.(Chopra*, et al.* 2018b). DNA was isolated from the candidate mutants using a Qiagen plant genomic DNA kit (Qiagen, Valencia, CA). DNA was subjected to WGS using an Illumina HiSeq 2500 sequencer (2x125 bp). Raw reads were analyzed using the method described by Chopra et al.(Chopra*, et al.* 2018b) to detect and predict the nature of mutations in each of these candidate mutants (Table **S3**). Gene-specific primers (Table **S6**) were designed to PCR amplify and sequence putative pennycress orthologs. PCR products were subjected to Sanger sequencing. DNA sequences were analyzed using the CLC Genomics Workbench (Qiagen, Valencia, CA) to process files, identify and confirm the corresponding mutations. Protein sequence alignments were performed using the sequences from *Thlaspi arvense* (Ta), and *Arabidopsis thaliana* (At) for the candidate genes. Mutation sites of the candidate genes were changed manually, and alignments were performed using the clustalW program. Genomic DNA sequences, protein sequences and alignments for all the wild-type and mutants genes discussed in this article are provided in the **Data S1, S2,** and **S3** files.

512 **Early flowering phenotyping**

513 F_2 plants from a cross between the wild-type GRIN accession Ames
514 23761 and the early flowering mutant MN A7-25 were planted into Ray Leach
515 SC10 Cone-tainers™ filled with Sun Gro Metro Mix 560 Sun-Coir. After
516 germination, plants were allowed to grow to the two true leaf stage in a 20 ºC
517 growth chamber. Plants were then transferred to a growth chamber maintained at
518 4 ºC and 8 hours of light for a vernalization period of 21 days. After this period
519 plants were returned to the 20 ºC growth chamber. Days to flowering were
520 recorded as the number of days after plants were returned to the 20 ºC growth
521 chamber to the first open florets visible on the plant.
522
523 **Genotyping using allele-specific markers**
524 To perform co-segregation analysis on F_2 progenies derived from *elf6-*
525 *1**Ames23761 and *aop2-like-1**MN106 crosses, we designed allele-specific and
526 flanking primers for each of the alleles (Table **S4**). DNA was extracted using the
527 Sigma ready extract method and genotyping was performed using the methods
528 described in Chopra et al. (Chopra*, et al.* 2018b).
529
530 **Expression analysis**
531 Leaf tissue from ten seedlings for each replicate were pooled from wild-
532 type and the mutant (*elf6-1*) for RNA extractions. RNA was extracted using the
533 RNAeasy mini clean up kit (Qiagen, Valencia, CA) and treated with turbo DNase
534 (ThermoFisher Cat. No. AM2238). To evaluate the expression patterns in the
535 wild type and mutants; qRT-PCR primers were designed for actin, tubulin, *ELF6*
536 and *FLC* genes (Table **S7**). Briefly, cDNA libraries were synthesized using
537 Invitrogen cDNA synthesis kit (Invitrogen, Grand Island, NY) and PCR was
538 performed using SybrGreen (Roche Cat. No. 04 707 516 001) on a LightCycler
539 480 (Roche, Basel Switzerland). Average C_t values generated from three
540 replicates for each of the cDNA libraries were used to normalize and calculate
541 the fold change in expression of genes, respectively.
542

Reduced pod shatter phenotyping

The force required to break apart seedpods at the septum was determined using a gram force tension gauge (SSEYL ATG-100-2 Tension Gauge) attached to a two-inch alligator clip (Gardner Bender, Milwaukee, WI). Briefly, one side of a pod was clipped and then the other side was manually pulled until either the pod broke at the septum or the fin tore. Pods that broke on handling before being clipped were recorded as zero. For each line, ten pods each from five different plants were used for the measurements.

Samples for observations of separation layers were prepared by either free-hand or microtome sectioning. The free-hand sections were cut with a razor blade, briefly stained with 0.05% toluidine blue O, and then transferred to a microscope slide for observation using Nikon SMZ1500 stereo microscope. Samples used for microtome sectioning were fixed for 5 h in 2% glutaraldehyde buffered with 0.025M phosphate buffer, pH7, and post-fixed 1–2 h in 2% OsO4. Specimens were dehydrated in a graded series of acetone, infiltrated in Spurr's resin, and polymerized in a 70° C oven. Semi-thin sections 1 µm thick were made with a Leica UCT Ultracut Microtome, stained with 0.1% toluidine blue, and viewed with a Leica DMRBE compound microscope using the 20x objective.

Glucosinolates in leaf tissue and seeds using UV absorbance method

To determine glucosinolate content in the pennycress leaf tissues and pennycress seeds, we recorded the fresh tissue weights or seed weights of the samples. Briefly, glucosinolates were extracted in 80% methanol followed by purification and were quantified using the method described by Chopra et al. (Chopra*, et al.* 2018a) . At least three biological replicates were used for estimating the glucosinolate content in each of the segregating progeny.

HPLC analysis of wild-type and Nutty seeds

Glucosinolates in the seeds of the wild-type and mutant samples were extracted and analyzed with HPLC using the method described by Kliebenstein et al. (Kliebenstein*, et al.* 2001). Briefly, forty microliters of the glucosinolate

574 extract was run on a 5-mm column (Lichrocart 250–4 RP18e, Hewlett-Packard,
575 Waldbronn, Germany) on a Hewlett-Packard 1100 series HPLC. Compounds
576 were detected at 229 nm and separated utilizing the programs described by
577 Kliebenstein et al.(Kliebenstein*, et al.* 2001) with aqueous acetonitrile.
578
579 **Fatty acid composition analysis using Gas Chromatography**
580 Approximately 100 mg of seeds from the wild-type and mutants of interest
581 from NIRS scans were weighed and crushed with 1,000 µl of hexane containing
582 a C17:0 internal standard using a mechanical homogenizer for three min at 10
583 m/s. The hexane supernatant containing the extracted oil was then transferred to
584 glass vials for methylation and FAMES were separated and detected using the
585 methods described in Chopra et al (Chopra*, et al.* 2018a). At least two replicates
586 were used for estimating the fatty acid composition in each of the lines.
587

Acknowledgments

General: We acknowledge the hard work of many undergraduates and others who contributed to this study including: Karl Nord, Carl Branch, Liam Sullivan, Luke Aldrich, Greta Rockstad and many others. We thank Martha Cook and Tara Nazarenus for help with tissue embedding / sectioning / microscopy and seed oil analyses, respectively.

Funding: This material is based upon work that is supported by the National Institute of Food and Agriculture, U.S. Department of Agriculture, under award numbers 2014-67009-22305 and 2018-67009-27374 to M.D.M, J.C.S and W.B.P. Funding for research conducted in the E.B.C lab was provided by U.S. Department of Energy, Office of Science, OBER (DOE-BER SC0012459) and NSF Plant Genome Program (13-39385). Additional funds were provided by the Minnesota Department of Agriculture and the University of Minnesota Forever Green Initiative to J.A.A. and B.I., and the Walton Family Foundation, PepsiCo and General Mills to D.L.W.

Author Contributions M.D.M. and J.C.S. conceived the study, designed experiments, supervised and organized coworkers, created mutagenized populations, isolated mutants, helped characterize the mutants, identified candidate genes, and co-wrote the manuscript. R.C. along with E.B.C, C.G. and B.I. characterized the fatty acid profiles, along with K.M.D. performed WGS analyses, and also independently performed extensive data analyses and helped write the first draft of the manuscript. J.L. D.J.K, and T.U. were responsible for the wet lab glucosinolate analyses. E.D., N.F, R.E, M.E., M.M., A.B. and K.A. isolated mutants and helped characterize candidate genes. M.O. helped evaluate *FLC* expression and seed yield of the *elf6* mutant. J.A.A. and K.A. help with the initial NIR set-up. K.M.D, K.F. and R.C. helped characterize the *elf6* mutant. A.B. generated and imaged wild-type and *ind-3* seedpod sections. D.L.W. isolated wild-type line MN106 used in the study and aided in planning.

Competing Interests: The authors declare potential competing financial interests as intellectual property applications have been submitted on portions of the reported research.

622 **Data and materials availability:** All sequence information described in this study is
623 contained within the Supplementary Materials. All plant materials described in this report
624 are available upon completion of Material Transfer Agreements.
625
626

627
628 Literature Cited
629
630 (2014) *Yield Gains in Major U.S. Field Crops* Madison, WI: American Society of Agronomy, Inc., Crop
631 Science Society of America, Inc., and Soil Science Society of America, Inc.
632 **Abbo, S., Pinhasi van-Oss, R., Gopher, A., Saranga, Y., Ofner, I. and Peleg, Z.** (2014) Plant
633 domestication versus crop evolution: a conceptual framework for cereals and grain legumes.
634 *Trends Plant Sci*, **19**, 351-360.
635 **Ahloowalia, B.S., Maluszynski, M. and Nichterlein, K.** (2004) Global impact of mutation-derived
636 varieties. *Euphytica*, **135**, 187-204.
637 **Anderson, T.R., Hawkins, E. and Jones, P.D.** (2016) CO2, the greenhouse effect and global warming:
638 from the pioneering work of Arrhenius and Callendar to today's Earth System Models. *Endeavour*,
639 **40**, 178-187.
640 **Bell, J.** (1982) From rapeseed to canola: A brief history of research for superior meal and edible oil.
641 *Poultry Sci*, **61**, 613-622.
642 **Butler, L.G.** (1992) Antinutritional Effects of Condensed and Hydrolyzable Tannins. In *Plant*
643 *Polyphenols: Synthesis, Properties, Significance* (Hemingway, R.W. and Laks, P.E. eds). Boston,
644 MA: Springer US, pp. 693-698.
645 **Chopra, R., Folstad, N., Lyons, J., Ulmasov, T., Gallaher, C., Sullivan, L., McGovern, A., Mitacek, R.,**
646 **Frels, K., Altendorf, K., Killiam, A., Ismail, B., Anderson, J.A., Wyse, D.L. and Marks, M.D.**
647 (2018a) The Adaptable Use of Brassica NIRS Calibration Equations to Identify Pennycress
648 Variants to Facilitate the Rapid Domestication of a New Winter Oilseed Crop. *in press*.
649 **Chopra, R., Johnson, E.B., Daniels, E., McGinn, M., Dorn, K.M., Esfahanian, M., Folstad, N.,**
650 **Amundson, K., Altendorf, K., Betts, K., Frels, K., Anderson, J.A., Wyse, D.L., Sedbrook, J.C.**
651 **and Marks, M.D.** (2018b) Translational genomics using Arabidopsis as a model enables the
652 characterization of pennycress genes through forward and reverse genetics. *The Plant Journal*, **96**,
653 1093-1105.
654 **Claver, A., Rey, R., López, M.V., Picorel, R. and Alfonso, M.** (2017) Identification of target genes and
655 processes involved in erucic acid accumulation during seed development in the biodiesel feedstock
656 Pennycress (Thlaspi arvense L.). *J Plant Physiol*, **208**, 7-16.
657 **CrevillÃ©n, P., Yang, H., Cui, X., Greeff, C., Trick, M., Qiu, Q., Cao, X. and Dean, C.** (2014)
658 Epigenetic reprogramming that prevents transgenerational inheritance of the vernalized state.
659 *Nature*, **515**, 587.
660 **De Bruin, J.L. and Pedersen, P.** (2008) Soybean Seed Yield Response to Planting Date and Seeding Rate
661 in the Upper Midwest All rights reserved. No part of this periodical may be reproduced or
662 transmitted in any form or by any means, electronic or mechanical, including photocopying,
663 recording, or any information storage and retrieval system, without permission in writing from the
664 publisher. *Agron J*, **100**, 696-703.
665 **DeHaan, L.R., Van Tassel, D.L., Anderson, J.A., Asselin, S.R., Barnes, R., Baute, G.J., Cattani, D.J.,**
666 **Culman, S.W., Dorn, K.M. and Hulke, B.S.** (2016) A pipeline strategy for grain crop
667 domestication. *Crop Sci*, **56**, 917-930.
668 **Dorn, K.M., Fankhauser, J.D., Wyse, D.L. and Marks, M.D.** (2013) De novo assembly of the
669 pennycress (Thlaspi arvense) transcriptome provides tools for the development of a winter cover
670 crop and biodiesel feedstock. *Plant Journal*, **75**, 1028-1038.
671 **Dorn, K.M., Fankhauser, J.D., Wyse, D.L. and Marks, M.D.** (2015) A draft genome of field pennycress
672 (Thlaspi arvense) provides tools for the domestication of a new winter biofuel crop. *DNA*
673 *Research*, **22**, 121-131.
674 **Dorn, K.M., Johnson, E.B., Daniels, E., Wyse, D. and Marks, M.D.** (2017) Spring flowering habit in
675 field pennycress (Thlaspi arvense) has arisen multiple independent times. *Plant Direct*, **2**, e00097.
676 **Dunn, M., Ulrich-Schad, J.D., Prokopy, L.S., Myers, R.L., Watts, C.R. and Scanlon, K.** (2016)
677 Perceptions and use of cover crops among early adopters: Findings from a national survey.
678 *Journal of Soil and Water Conservation*, **71**, 29-40.
679 **Dyer, J.M., Stymne, S., Green, A.G. and Carlsson, A.S.** (2008) High-value oils from plants. *The Plant*
680 *Journal*, **54**, 640-655.
681 **Fan, J., Shonnard, D.R., Kalnes, T.N., Johnsen, P.B. and Rao, S.** (2013) A life cycle assessment of
682 pennycress (Thlaspi arvense L.)-derived jet fuel and diesel. *Biomass and Bioenergy*, **55**, 87-100.

683 **Gerland, P., Raftery, A.E., Ševčíková, H., Li, N., Gu, D., Spoorenberg, T., Alkema, L., Fosdick, B.K., Chunn, J., Lalic, N., Bay, G., Buettner, T., Heilig, G.K. and Wilmoth, J.** (2014) World population stabilization unlikely this century. *Science*, **346**, 234-237.

686 **Gordon, M.H.** (2001) The development of oxidative rancidity in foods. In *Antioxidants in food*: Elsevier, pp. 7-21.

688 **Hart, G.E.** (2015) Feeding the ethanol boom: Where will the corn come from. *Iowa Ag Review*, **12**.

689 **Harvey, M. and Pilgrim, S.** (2011) The new competition for land: Food, energy, and climate change. *Food Policy*, **36**, S40-S51.

691 **Heap, I.** (2014) Herbicide Resistant Weeds. In *Integrated Pest Management: Pesticide Problems, Vol.3* (Pimentel, D. and Peshin, R. eds). Dordrecht: Springer Netherlands, pp. 281-301.

693 **Ho, D.P., Ngo, H.H. and Guo, W.** (2014) A mini review on renewable sources for biofuel. *Bioresource Technology*, **169**, 742-749.

695 **Isbell, T.A.** (2009) US effort in the development of new crops (Lesquerella, Pennycress Coriander and Cuphea). *Oléagineux, Corps gras, Lipides*, **16**, 205-210.

697 **James, D.W., Jr., Lim, E., Keller, J., Plooy, I., Ralston, E. and Dooner, H.K.** (1995) Directed tagging of the Arabidopsis FATTY ACID ELONGATION1 (FAE1) gene with the maize transposon activator. *Plant Cell*, **7**, 309-319.

700 **Jeong, J.H., Song, H.R., Ko, J.H., Jeong, Y.M., Kwon, Y.E., Seol, J.H., Amasino, R.M., Noh, B. and Noh, Y.S.** (2009) Repression of FLOWERING LOCUS T chromatin by functionally redundant histone H3 lysine 4 demethylases in Arabidopsis. *PLoS One*, **4**, e8033.

703 **Johnson, G.A., Kantar, K.B., Betts, K.J. and Wyse, D.L.** (2015) Field Pennycress Production and Weed Control in a Double Crop System with Soybean in Minnesota. *Agron J*, **107**, 532-540.

705 **Johnson, G.A., Wells, M.S., Anderson, K., Gesch, R.W., Forcella, F. and Wyse, D.L.** (2017) Yield Tradeoffs and Nitrogen between Pennycress, Camelina, and Soybean in Relay- and Double-Crop Systems. *Agron J*, **109**, 2128-2135.

708 **Jordan, N., Boody, G., Broussard, W., Glover, J.D., Keeney, D., McCown, B.H., McIsaac, G., Muller, M., Murray, H., Neal, J., Pansing, C., Turner, R.E., Warner, K. and Wyse, D.** (2007) Environment - Sustainable development of the agricultural bio-economy. *Science*, **316**, 1570-1571.

711 **Kliebenstein, D.J., Lambrix Vm Fau - Reichelt, M., Reichelt M Fau - Gershenzon, J., Gershenzon J Fau - Mitchell-Olds, T. and Mitchell-Olds, T.** (2001) Gene duplication in the diversification of secondary metabolism: tandem 2-oxoglutarate-dependent dioxygenases control glucosinolate biosynthesis in Arabidopsis. *Plant Cell*, **13 (3)**, 681-693.

715 **Knutsen, H.K., Alexander, J., BarregAard, L., Bignami, M., Brüschweiler, B., Ceccatelli, S., Dinovi, M., Edler, L., Grasl‐Kraupp, B. and Hogstrand, C.** (2016) Erucic acid in feed and food. *EFSA Journal*, **14**, 1-173.

718 **Liljegren, S.J., Roeder, A.H., Kempin, S.A., Gremski, K., Ostergaard, L., Guimil, S., Reyes, D.K. and Yanofsky, M.F.** (2004) Control of fruit patterning in Arabidopsis by INDEHISCENT. *Cell*, **116**, 843-853.

721 **Lu, C., Xin, Z., Ren, Z., Miquel, M. and Browse, J.** (2009) An enzyme regulating triacylglycerol composition is encoded by the ROD1 gene of Arabidopsis. *Proc Natl Acad Sci U S A*, **106**, 18837-18842.

724 **Lu, F., Li, G., Cui, X., Liu, C., Wang, X.-J. and Cao, X.** (2008) Comparative Analysis of JmjC Domain-containing Proteins Reveals the Potential Histone Demethylases in Arabidopsis and Rice. *J Integr Plant Biol*, **50**, 886-896.

727 **McGinn, M., Phippen, W.B., Chopra, R., Bansal, S., Jarvis, B.A., Phippen, M.E., Dorn, K.M., Esfahanian, M., Nazarenus, T.J., Cahoon, E.B., Durrett, T.P., Marks, M.D. and Sedbrook, J.C.** (2018) Molecular tools enabling pennycress (Thlaspi arvense) as a model plant and oilseed cash cover crop. *Plant Biotechnol J*, **0**.

731 **Moser, B.R.** (2012) Biodiesel from alternative oilseed feedstocks: camelina and field pennycress. *Biofuels*, **3**, 193-209.

733 **Moser, B.R., Knothe, G., Vaughn, S.F. and Isbell, T.A.** (2009a) Production and Evaluation of Biodiesel from Field Pennycress (Thlaspi arvense L.) Oil. *Energ Fuel*, **23**, 4149-4155.

735 **Moser, B.R., Shah, S.N., Winkler-Moser, J.K., Vaughn, S.F. and Evangelista, R.L.** (2009b) Composition and physical properties of cress (Lepidium sativum L.) and field pennycress (Thlaspi arvense L.) oils. *Ind Crop Prod*, **30**, 199-205.

738 **Noh, B., Lee, S.H., Kim, H.J., Yi, G., Shin, E.A., Lee, M., Jung, K.J., Doyle, M.R., Amasino, R.M. and**
739 **Noh, Y.S.** (2004) Divergent roles of a pair of homologous jumonji/zinc-finger-class transcription
740 factor proteins in the regulation of Arabidopsis flowering time. *Plant Cell*, **16**, 2601-2613.
741 **Oladosu, Y., Rafii, M.Y., Abdullah, N., Hussin, G., Ramli, A., Rahim, H.A., Miah, G. and Usman, M.**
742 (2016) Principle and application of plant mutagenesis in crop improvement: a review.
743 *Biotechnology & Biotechnological Equipment*, **30**, 1-16.
744 **Ott, M.A., Eberle, C.D., Thom, M.D., Archer, D.W., Forcella, F., Gesch, R.W. and Wyse, D.L.** (2019)
745 Economics and agronomics of dual-cropping pennycress and camelina with soybean in Minnesota.
746 *Agron J*, **in press**.
747 **Perng, W., Villamor, E., Mora-Plazas, M., Marin, C. and Baylin, A.** (2014) Alpha-linolenic acid (ALA)
748 is inversely related to development of adiposity in school-age children. *European Journal Of*
749 *Clinical Nutrition*, **69**, 167.
750 **Phippen, W.B. and Phippen, M.E.** (2012) Soybean Seed Yield and Quality as a Response to Field
751 Pennycress Residue. *Crop Sci*, **52**, 2767-2773.
752 **Price, A.J., Balkcom, K.S., Culpepper, S.A., Kelton, J.A., Nichols, R.L. and Schomberg, H.** (2011)
753 Glyphosate-resistant Palmer amaranth: A threat to conservation tillage. *Journal of Soil and Water*
754 *Conservation*, **66**, 265-275.
755 **Provart, N.J., Alonso, J., Assmann, S.M., Bergmann, D., Brady, S.M., Brkljacic, J., Browse, J.,**
756 **Chapple, C., Colot, V., Cutler, S., Dangl, J., Ehrhardt, D., Friesner, J.D., Frommer, W.B.,**
757 **Grotewold, E., Meyerowitz, E., Nemhauser, J., Nordborg, M., Pikaard, C., Shanklin, J.,**
758 **Somerville, C., Stitt, M., Torii, K.U., Waese, J., Wagner, D. and McCourt, P.** (2015) 50 years
759 of Arabidopsis research: highlights and future directions. *New Phytol.*, **209**, 921-944.
760 **Rajaram, S.** (2014) Health benefits of plant-derived Î±-linolenic acid. *Am J Clin Nutr*, **100**, 443S-448S.
761 **Robertson, G.P. and Vitousek, P.M.** (2009) Nitrogen in Agriculture: Balancing the Cost of an Essential
762 Resource. *Annual Review of Environment and Resources*, **34**, 97-125.
763 **Roesch-McNally, G.E., Basche, A.D., Arbuckle, J.G., Tyndall, J.C., Miguez, F.E., Bowman, T. and**
764 **Clay, R.** (2017) The trouble with cover crops: Farmers' experiences with overcoming barriers to
765 adoption. *Renew Agr Food Syst*, **33**, 322-333.
766 **Sedbrook, J.C., Phippen, W.B. and Marks, M.D.** (2014) New approaches to facilitate rapid
767 domestication of a wild plant to an oilseed crop: example pennycress (Thlaspi arvense L.). *Plant*
768 *Sci*, **227**, 122-132.
769 **Sonderby, I.E., Geu-Flores, F. and Halkier, B.A.** (2010) Biosynthesis of glucosinolates--gene discovery
770 and beyond. *Trends Plant Sci*, **15**, 283-290.
771 **Tester, M. and Langridge, P.** (2010) Breeding Technologies to Increase Crop Production in a Changing
772 World. *Science*, **327**, 818.
773 **Tilman, D., Balzer, C., Hill, J. and Befort, B.L.** (2011) Global food demand and the sustainable
774 intensification of agriculture. *Proceedings of the National Academy of Sciences*, **108**, 20260.
775 **Warwick, S.I., Francis, A. and Susko, D.J.** (2002) The biology of Canadian weeds. 9. Thlaspi arvense L.
776 (updated). *Can J Plant Sci*, **82**, 803-823.
777 **Weyers, S., Thom, M.D., Forcella, F., Eberle, C.A., Matthees, H., Gesch, R., Ott, M.A., Feyereisen, G.,**
778 **Strock, J.S. and Wyse, D.** (201) Reduced-nutrient leachates in cash cover crop-soybean systems.
779 *J Environ Qual*, **in press**.
780 **Wittkop, B., Snowdon, R.J. and Friedt, W.** (2009) Status and perspectives of breeding for enhanced yield
781 and quality of oilseed crops for Europe. *Euphytica*, **170**, 131.
782 **Wortmann, C.S., Tarkalson, D.D., Shapiro, C.A., Dobermann, A.R., Ferguson, R.B., Hergert, G.W.**
783 **and Walters, D.** (2011) Nitrogen Use Efficiency of Irrigated Corn for Three Cropping Systems in
784 Nebraska. *Agron J*, **103**, 76-84.
785
786

Supplementary Materials

List

Figure S1. Comparisons of the separation layer of wild-type Spring 32 (left) and *ind-3* (right) seedpod sections stained with toluidine blue.

Figure S2. Average glucosinolate content of the progenies from the segregating population of *Ta-aop2-1* plants.

Figure S3. Comparisons of a wild-type pennycress plant with the "Nutty" *aop2-1* and "canola-like-oil" *fae1-1 rod1-1* mutants grown in controlled chambers.

Figure S4. A schematic representation of pennycress life-cycle from seeds to a mature plant.

Figure S5. F1 plant carrying all of the described mutations in this report and was selected using the allele-specific primers and strategy described in Fig S4.

Table S1. Plant height and seed yield comparisons between wild-type and early flowering mutant (elf6-1) grown in 1 x 1 m plots in the field.

Table S2. Protein domain analysis to highlight the conserved region in the Indehiscent gene (IND) using several databases.

Table S3. Summary of the mutations in the domestication related genes characterized in this study.

Table S4. Allele-specific primers designed for single nucleotide mutations in the domestication candidate genes.

Table S5. Summary of double and triple mutants generated using the approach described in Fig S4.

Table S6. Primers used for Sanger sequencing.

Table S7. Primers used for expression analysis of the *Ta-elf6-1* early flowering mutant.

Data S1. Genomic DNA Sequences of the candidate genes for pennycress domestication obtained from Thlaspi (pennycress) version 1.0 gene annotations and information on the mutation sites.

817 **Data S2.** Predicted protein sequences of the candidate genes for pennycress
818 domestication obtained from Thlaspi version 1.0 gene annotations and mutation
819 and information on the mutation sites.
820 **Data S3.** Protein sequence alignments of *Thlaspi arvense* (pennycress) wild-type
821 sequences with the corresponding mutant and orthologous Arabidopsis
822 sequences.
823
824
825
826
827
828
829
830
831
832
833
834

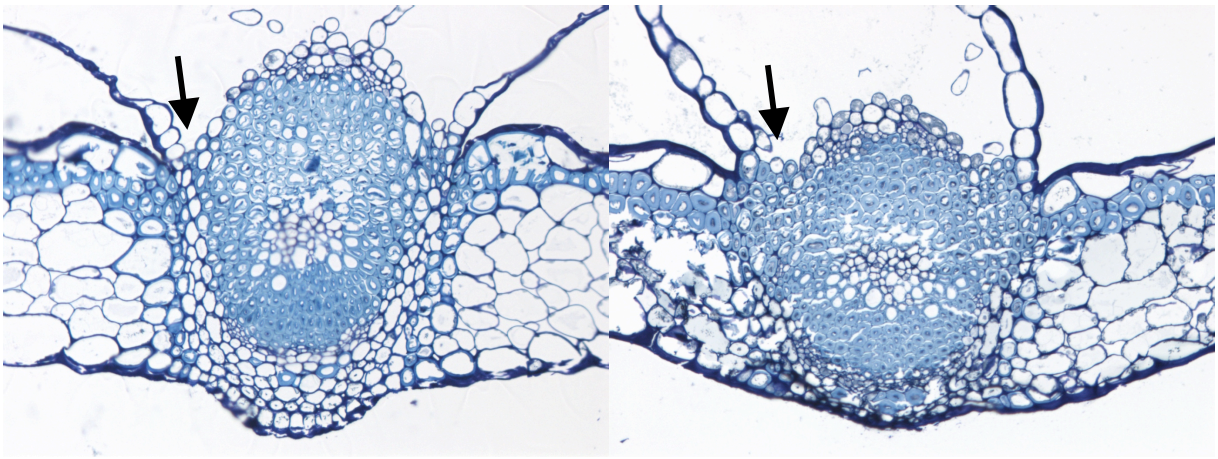


Figure S1. Comparisons of the separation layer of wild-type Spring 32 (left) and *ind-3* (right) seedpod sections stained with toluidine blue. Note the lack of a clear separation layer in *ind-3* (arrows).

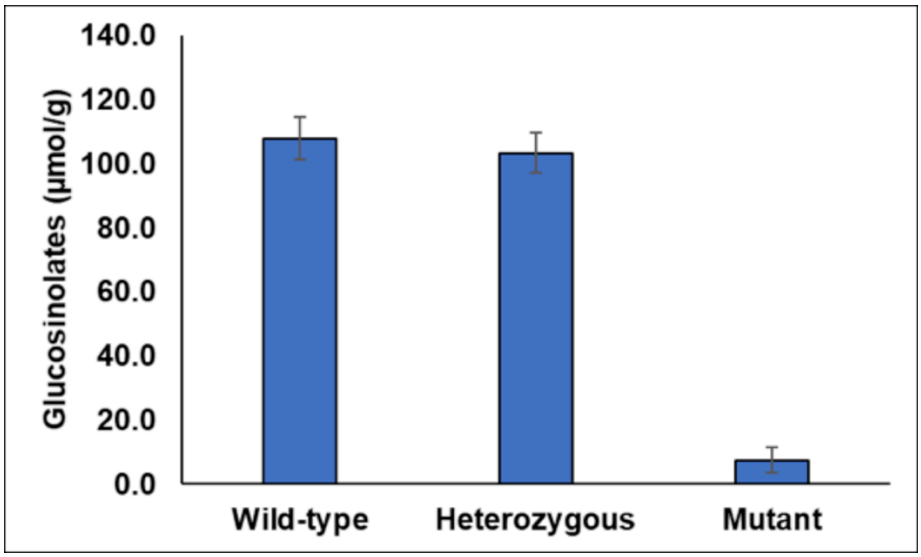
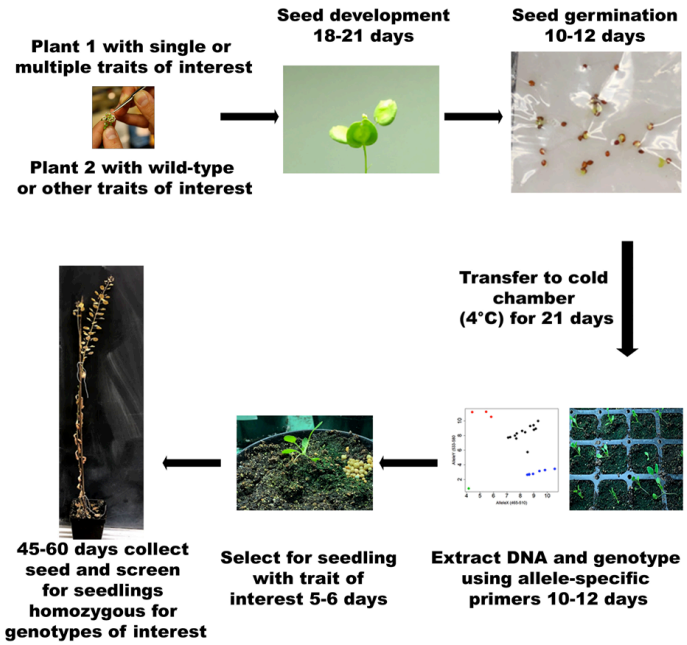


Figure S2: Average glucosinolate content of the progenies from the segregating population of *Ta-aop2-1* plants. Allele-specific markers were used to assess the *aop2-1* genotype of members in the population and glucosinolate content was analyzed using NIRS and confirmed with wet-lab assay.

Figure S3. Comparisons of a wild-type pennycress plant with the "Nutty" *aop2-1* and "canola-like-oil" *fae1-1 rod1-1* mutants grown in controlled chambers. No developmental defects were observed to be associated with the respective traits.



872

873 **Figure S4.** A schematic representation of pennycress life-cycle from seeds to a
874 mature plant. Seeds can be developed from the crossing event in 12-15 days.
875 Seeds can be propagated from seeds to mature in 75-80 days. This scheme of
876 propagating pennycress from would allow for minimum of three generations per
877 year.

878

879

880

881 **Figure S5.** F_1 plant carrying all of the described mutations in this report and was
882 selected using the allele-specific primers and strategy described in Fig S4.
883

884

885 **Table S1:** Plant height and seed yield comparisons between wild-type and early

886 flowering mutant (elf6-1) grown in 1 x 1 m plots in the field. No statistically

887 significant differences were observed among these lines (± sd).

888

Line	Plant height (cm)	Seed yield (g)
Wild-type	65 (±0.25)	46.56 (±11.48)
elf6-1	67.58 (±0.76)	57.61 (±14.15)

889

890

Table S2. Protein domain analysis to highlight the conserved region in the Indehiscent gene (IND) using several databases.

Gene	Database	Family number	Start Pos.	End Pos.	Gene Family
Ta1.0_25465	SMART	SM00353	98	147	Myc-type, basic helix-loop-helix (bHLH) domain
Ta1.0_25465	Gene3D	G3DSA:4.10.280.10	84	154	Helix-loop-helix DNA-binding domain superfamily
Ta1.0_25465	ProSiteProfiles	PS50888	92	141	Myc-type, basic helix-loop-helix (bHLH) domain
Ta1.0_25465	SUPERFAMILY	SSF47459	97	151	Helix-loop-helix DNA-binding domain superfamily
Ta1.0_25465	CDD	cd00083	96	146	Myc-type, basic helix-loop-helix (bHLH) domain
Ta1.0_25465	Pfam	PF00010	100	141	Myc-type, basic helix-loop-helix (bHLH) domain

896 **Table S3.** Summary of the mutations in the domestication related genes
897 characterized in this study.
898
899
900

Scaffold Number	Gene	# of Alleles	Mutation
Ta_scaffold_10	*ELF6* Early Flowering6	1	**Mutant:** A7 25 **Genomic pos:** 107096 **cDNA pos:** C952T **AA pos:** His318Tyr
Ta_scaffold_1003	*IND* Indehiscent	3	**Mutants:** E5 552; E5 550; Spring 32 NS **Genomic pos**: 6741, 6754, 6795 **cDNA pos**: C247T; G260A; G301A **AA pos**: Pro83Ser; Arg87His; Ala101Thr
Ta_scaffold_74	*AOP2* Alkenyl- and Hydroxyalkyl-Producing2	1	**Mutant**: E3196 **Genomic pos**: 395125 **cDNA pos**: G289A **AA pos**: Gly97Arg
Ta_scaffold_45	*FAE1* Fatty Acid Elongation1	2	**Mutants**: V296; K1822 **Genomic pos**: 254962; **cDNA pos**: C1018T; G1349A **AA pos**: Gln340Stop; Trp450Stop
Ta_scaffold_199	*ROD1* Reduced Oleate Desaturation1	2	**Mutants**: d0422; E5 370 P6 **Genomic pos**: 186474; 186409 **cDNA pos**: G678A; G743A **AA pos**: Met226Ile; Arg248Lys

Table S4. Allele-specific primers designed for single nucleotide mutations in the domestication candidate genes.

Trait	Primer Name	Sequence
Early Flowering	elf6-1_C	ATCCCTGGTGTGACATCTCC
	elf6-1_X	AAGCTCATGGTCCTCAACGTG
	elf6-1_Y	AAGCTCATGGTCCTCAACGTA
Reduced Shatter	ind-1_C	TCGTCGCTTACCCTTACGTT
	ind-1_X	CGATCCAGCCACCGTTC
	ind-1_Y	CGATCCAGCCACCGTTT
	ind-2_C	TCTGAGTGTCGTCGCTTACC
	ind-2_X	CCGTTCCTAAACCGAACCG
	ind-2_Y	CCGTTCCTAAACCGAACCA
	ind-3_C	CTGGCACCATCCTCTTCAAT
	ind-3_X	GACACTCAGACGGTGGTGG
	ind-3_Y	GACACTCAGACGGTGGTGA
Reduced Glucosinolates	aop2_C	CGTCACAATGATCAGGACGTA
	aop2_X	CGGTCTTTCCGAGAGTATGG
	aop2_Y	AACGGTCTTTCCGAGAGTATGA
Reduced Erucic Acid	fae1-1_C	TAAAACCGGGGTGTGTTTGT
	fae1-1_X	ACCCAATGTTGTTATGTTTTTCTG
	fae1-1_Y	GACCCAATGTTGTTATGTTTTTCTA
	fae1-2_C	AAACCCTGACCCTAAAGCAA
	fae1-2_X	ATGAAGAGAGGGAACAAAGTGTG
	fae1-2_Y	GATGAAGAGAGGGAACAAAGTGTA
Reduced polyunsaturated fatty acids	rod1-1_C	GGTTCGATGATCGCATCTTT
	rod1-1_X	CTCATCCTCTGCATTCTCCTC
	rod1-1_Y	TCTCATCCTCTGCATTCTCCTT
	rod1-2_C	GCGATGCTTTTTGACATCCT
	rod1-2_X	CTCGTCCCGAGCAGCC
	rod1-3_Y	CTCGTCCCGAGCAGCT

907 **Table S5:** Summary of double and triple mutants generated using the approach
908 described in Fig S4.
909

Combination of alleles
fae1-1/rod1-1
fae1-1/elf6-1
fae1-1/aop2-like-1
fae1-1/ind1-1
rod1-1/aop2-like-1
elf6-1/ind1-1
aop2-like-1/ind1-1
fae1-1/rod1-1/aop2-like-1
fae1-1/rod1-1/ind1-1

910
911

912 **Table S6.** Primers used for Sanger sequencing.
913

Trait	Gene	Forward Primer	Reverse Primer
Early Flowering	*ELF6*	CGTCCAAGCAGAAGAACATGG	CCAGTTGATCAATGTTTCTGC
Reduced Shatter	*IND1*	TGAAGGAGATGCAGTACACG	TATAACGGATGGCTTCGTCG
Reduced Glucosinolates	*AOP2*	TCGGAGGAGCTTAAGAAGTC	ACCTGATGCTCTTGTTACCG
Reduced Erucic Acid	*FAE1-1*	GACCTAAGTTCTTCTGTAGC	TAAAACCGGGGTGTGTTTGT
Reduced Erucic Acid	*FAE1-2*	GAGAAAACATCGTAGCCATC	AAACCCTGACCCTAAAGCAA
Reduced PUFAs	*ROD1*	CATGTGGGGTTTGGGTTAAC	GTTCAAGTAATTAACAGTATATTC

914
915

916 **Table S7.** Primers used for expression analysis of the *Ta-elf6-1* early flowering
917 mutant.
918

Gene	Forward Primers	Reverse Primers
ELF6-qPCR	CCTGGTGAATTTGTTGTGA	GGACAGCATGGGAAGATA
FLC-qPCR	GCTATCAACAAGCTTC	GCACCATGAGCTACTA
Actin-qPCR	GTGAGACACACCATCACCAGAAT	TGTCGCCATCCAAGCTGTTCT
Ubiquitin-qPCR	AGTTAAGAGGACTGTCTGG	TCCTGAACCATATCCTCT

919
920

Data S1. Genomic DNA Sequences of the candidate genes for pennycress domestication obtained from Thlaspi (pennycress) version 1.0 gene annotations and information on the mutation sites. (Exons – bold; Introns – small letters).

Ta ELF6

***Ta-elf6-1* (highlighted red; substitution: C to T; cDNA pos:952)**

ATGGGTGATGTTGAAATTCCCAATTGGCTAAAAGCCTTGCCTTTGGCACCTGTCTTTAGACCTACGGACACCGAATTCG
CAGATCCTATCGCGTATATATCGAAAATCGAGAAAGAGGCCAGTGCTTTTGGGATCTGCAAGATCATTCCTCCTTTACC
CAAGCCGTCGAAAAAGTATGTTTTCTACAACTTGAACAAGTCTCTTTTGAGGTGTCCTGAATTGGCTTCGGATGTAGAC
ATTTCGAAAGTGTGTCAAGAGGATAGAGCTGTGTTCACCACTAGGCAGCAAGAGTTAGGGCAGGCTGTAAAACGAAAGA
AAGGAGGAGAGAGCAGTAAGAGCAATTCTCAAAGGAGTGGCGTTAAGCAGGTGTGGCAAAGTGGAGGCGTGTACACGTT
GGAGCAGTTCGAATCTAAGTCAAAAACTTTCTACAAAAGCCAGTTAGGAACCACAAAAGAAGTGCCACCGGTTGTGGTT
GAGGCATTGTTCTGGAAAGCAGCTTTAGAGAAGCCTATATACATAGAGTATGCAAATGATGTGCCTGGCTCGGCTTTCG
GTGAACCAGAGGGTCATTTCAGGCATTTTCGGCAGAGAAAGAGGAGAGGGAGAGGATTTTATCAGAGGAAGGCAGAGGT
CAGTGAAGACAGCGGAGTAGAAAACGGGACGAATAGTCAAGAACCAACCTGCAAGAATGGTGAGAAAACATTGCCTGAG
GTAGCAAAGGCATCTCTTGCTTCTCCGAGTTTATTATCTCAGGATCCGTCCAAGCAGAAGAACATGGATATTGTTGATG
AAATGGAAGGTACTGCAGGCTGGAAGCTCTCCAACAGTTCATGGAACCTTCAGATGATTGCACGTTCACCTGGATCTGT
TACACGCTTCATGCCAGATGACATCCCTGGTGTCACATCTCCCATGGTTTATATCGGTATGTTGTTCAGCTGGTTTGCC
TGGCACGTTGAGGACCATGAGCTTCACAGTATGAATTACCTTCACACTGGCTCGCCAAAGACGTGGTACGCTGTCCCTG
GTGATTATGCATTTGACTTTGAAGAGGTTATCCGCAAAAATTCGTATGGCAGAAACATTGATCAACTGGgtacgttctt
tctgaaaagtactgctaaatatgatatactgtttctgtttatatagaaatgtttcgttggtgtaatacatcatacatgt
gagaaatgagatttcctagaatgattaccgcatccatatttttctttactagcacctttttttttgctttgtaagtga
aatgtggctgacattgactatgatatgacgagagtttgtactcttgggaaattgcgttaggacttattgctttaaggtt
attatgatagatatgagacgttgcaacacttcttatgaaatgcattgtccttctgtttctcattgactcttagctgttc
tttgtcactttcag**CTGCTCTCACCCAACTAGGCGAAAAGACAACTCTTGTATCACCTGAGATGATAATTGCATCTGAC**
ATTCCCTGCTGTAGgtaggccttttaattttatttgaactttcacttctgttatgtggagatgtgaggcagtttgtgtt
ttcttataactacgccaagctctgctatatctatttttgttttcccacgtag**GTTGGTACAGAATCCTGGTGAATTTGT**
TGTGACTTTTCCGAGGTCTTATCATGTAGGATTCAGCCACGgtaaaaatgctttttttcttcaaacattcttaagtctt
tgtgactttacttttggtcgtcccattttgcactcttcaaagtgtgtgagaaaatgtgaaaattcaaaattcaaaattga
gtaaagctttggagaaaaatgagtgttttacgacagagcataaggtgaggattgatcttctaattaggagaatgaagaa
ccaaatttctattaagtagtagttatataagttgcatagtaaaagcggatagtttggcttcgattaggaatacaaattg
caatatttttttcagaatccttaactaagcagaattaatttaacgttttaaag**GTTTTAACTGTGGGGAAGCCGCTAAT**
TTTGGAACTCCACAATGGCTCAACGTAGCTAAGGAAGCTGCTGTGCGACGGGCAGCCATGAATTATCTTCCCATGCTGT
CCCATCAGCAGCTGCTATATCTCTTGACCATGTCCTTTGTTTCAAGGCAAATTTCCATGGCCTCTTTgtacatagaacc
ctttctgctggaacctgttaatcctcatattcttgtaaatattaaaattttcag**AGTGCCACGATCATTACTACCAGG**
TGGTCGTAGCTCCCGACTGAGAGATCGTCAGAGAGAAGAAAGGGAGTTCCTTGTGAAAAAAGCTTTTGTAGAAGATATA
CTGAACGAAAACAAGAATTTATCTGTTCTTCATCGAGAACCGGGATTTCGTTTGGTGATGTGGGACCCTGATTTACTCC
CGCGTCATAGTGTACATGGTCTTGTAACTGTTGGGGGTGCTGCTGTTTCATCTCCAGCAGAGGGAAAAAATGAACTTGA
GGAGAAGAATAAAGAGAAGACTACTCTTTTAGAGGAATTGAGTTTGTTCATGGAGAAGCTGAAAGATGTATACTACGAC
GATGATGATGGTCTGCTTAATGATTTCCAGGTTGATTCTGGAACCTTGGCATGTGTGGCGTGTGGCGTTCTTGGCTTCC
CCTTTATGTCTGTGGTACAGCCTTCTGAAAATGCATTAAATGATCTTTCAGAGAGACGAGGAGAGATAGgtaacagacc
ctcatttttttaaccaaactatgaactacaccatcttcgtttgaagcctgttaattgtgcttctatctattctacag**AT**
GGTCAGGAAATTACGGCACTGTTGTCAGAAAAGTCTGACTGTGAATGGAACATGTCCTCCAGGTATATAAGACCTCGCA
TTTTCTGCCTCGAACACACTATTGAGCTCCAGAGACTGCTGGAGTCACGAGGTGGACTGAAGTTCCTTGTAATTTGCCA
TAAAGgtaagtacgcgtcatttgctattaaattcgatgccaaagagaatattttgatcattctgcttttaacttttttt
ggaattgttgcag**ACTTTCAAAAATTTAAGGCATATGCGGCTATAGTGGCAGAGGAAGTTAAAGTCCCTTTCAGCTATG**
ATGATATCCTGTTAGAGAGTGCATCTAAAGAAGAGTTGAGTCTGATTGATCTTGCAATTGAAGATGAAGAAAACAACGA
ACATGGCGTAGACTGGACCTCAAAACTTGGTATCAATTTACGGTACTGTGTTAAAGTGAGGAAAAATTCCCCTTCTACG
AAAATTCAGCATGCACTGTCGCTAGGTGGCTTGTTCTCCGATACAAACCACATGCTAGATATGTCAACTATCAAATGGC
TGCAGAGAAAATCACGCTCAAAAGCCAAACCCAGTTGTACCTCAAGCTTCACACCTCGTGAACATCTTGAAGTAAAAGT
AGACAGAAAATTAGGGGAGAAGGAAAAAGTTGAATCCCAAGCCGGAAGAAAGGAAGAAAAGATCATCCAGTACTCGAGA
AAGAAAAAGTTGAAGCCCAAGCCTTCTGAAGAACGAAGTCAGGAACTAACTATCTCAGCTAAATCAGAAGATTTTGAAA
ACACATGCAACACACTTGCCAAAAGGTCACATCATCATGGGGCAATGCATTCTGATATGAACAATGAAATTGGAGATTT
TGGGAGGAATGGGGTATCCTTTTCAGAAAATCATTGTAGCTCACCTTTCACTGGGGCACGCGGACAAGAACATCCCAAG
ATCATTATCAAGTTTGGCTCAGCATTACATGGGAATATTACAAGCAGTTCTAGTTTGGTGAATGGAATCTCTGCTGACC
TAACTTCCGTAACCAGAGAGCACCAAGGACACTCTATGACCAGCAATAATAATGGGTCGAACTCAAGTAATCATGATGG
CCCAATAAAGCTGTCTGGTGAGCATGTCAGTGACGTGTCTGTACGTGATGTTGATGAAGCGGTTGAAATGAGCGACCAA

GAGTTCGAAGAACTGAGGTCTACCGTCACTAACATTGAGGAGGAACAGCAATCAGAGATGGTGAGACCAACCGCACTTC
AGGTGGAGGGAGAGGAATCTATGTGTACGAGAGAAATCTTGAGCTCTGAAGATATTATGCACACTGAGCAGCAGCAAGA
GCAAACTCAACTGGGTTTAGAAGTTCCTGAAACTGACATTGCCAGTGAGAACATAGTTGTGGACATGATCCATGATGAT
GAACCTCTGGCAACTAGGGATATATTAAGTTCAAGCAACGGTGATCAAGCTTCTTCAAATGGCTTGCAAGCTCTAGATA
ATGAACTTAGCATGGAGAGCGAAGTTGCAAGCTCAGAAAACACCGAGGTTATAGAGGCGTCGCCCAATTCTATTATGCG
AGAAGCAAATAAGAAGCGGAGAATAGAATCAGAGTCTGAGACAAATGATAATCCAGATGGTAGCATTGGTTTCATAAGG
AGTCCTTGTGAAGGGTTGAGGTCAAGGGGTAGGAGGAGAGTGACGCGTGAAGCTTCAGTCAGTCTCACTGAAACGAGCG
ATGAAGAGAAGAAACCCGCTGCGAAAAGGTTCAAGAAAACTCCAAAGACTCGCTCGGGGAGTCATCACCAAGAAGACTC
CACGACAAGTCACCACAACCGTTGTAACCTAGAGGGATGCAAGATGACTTTCAAGAGTAAAGCAGAGTTACAAGCTCAC
CAAAGAAACCGCTGCGCACATGAAGGGTGTGGAAAAAAATTCAGGGCTCACAAATATCTGGTGCTTCATCAACGTGTTC
ATAACGATGATAGACCTTTTGTGTGCTCTTGGAAAGGATGTTCCATGACTTTCAAATGGCCATGGGCGAGGACCGAGCA
TTTGCGTCTGCACACGGGAGAGCGACCATACAAATGCAAGGTCGATGGATGTGGAATGTCGTTTAGGTTTGTGTCGGAT
TACAGCCGCCATAGACGGAAAAAGGGGCATTATGTGACATAG

Ta IND

Ta-ind-1 **(highlighted red; substitution: C to T; cDNA pos:247)**
Ta-ind-2 **(highlighted red; substitution: G to A; cDNA pos:260)**
Ta-ind-3 **(highlighted red; substitution: G to A; cDNA pos:301)**

ATGAATTGGAACAAACCTAATGATCTCATCACACAAGAATACCCCTTTCTCCACGATCCTCATCTCATGATAGATCCAC
CTCCCGAAACCCTAAGTCATTTCCAGCCCCCGCCGACACTTTTCTCCGGTCACGGAGGGGAGGAAGAAGAAGAAGAAGA
TAATGAAGAGGAAGAGATGGATGCGATGAAGGAGATGCAGTACACGATCGCTGCCATGCAGCCCGTGGACATCGATCCA
GCCACCGTTCCTAAACCGAACCGCCGTAACGTAAGGGTAAGCGACGACACTCAGACGGTGGTGACTCGTCGGCGTCGAG
AAAAGATAAGCGAGAAGATCCGAATATTGAAGAGGATGGTGCCAGGCGGTGCGAAGATGGACACAGCCTCCATGCTCGA
CGAAGCCATCCGTTATACCAAGTTCTTGAAACGGCAGGTGAAGCTTCTTCAGCCTCACTCTCAGCTTGGAGCTCCTATG
TCTGACCCCTCTTGCCTTTGTTATTACCACAACTCCCAAACCTAA

Ta AOP2

Ta-aop2-1 **(highlighted red; substitution: G to A; cDNA pos:289)**

ATGGGTTCACTTTCAAACACTCCTCAGCTTCCAGTCATCTACCTCTCGGACCAAACCCTAAAACCAGGAAGCTCAAAGT
GGGTTGAAGTCAGGAGTGATGTCCGTAAAGCTCTTGAAGAGTACGGCGGTTTCGAGGTGTCGTACGATAGAGTGTCGGA
GGAGCTTAAGAAGTCGGTTTTGCAAGCCATGGAAGAGCTTTTCGCGTTACCAGTTGAGGCTAAACAGAGAAACGTCTCT
CCTAAACCCTTCAGCGGTTATTCCACTCATAACGGTCTTTCCGAGAGTATGGGGATCCAGGATCCTCATGTTTTGGACA
AAGTTTACGAGTTTACTCAACTTCTACGTCCTGATCATTGTGACGGTAACAAGAGCATCAGgtaatttgtgaaaaatac
tcaatattgcttcataatataaaaatactcaatattgcttcctaatctttttggcagtttatttcactacataaaataa
acccgcttttacatttttattgtttgtggtataagaatattagttcactcaaacagcatgaaactaataattgaaattt
tgtatttgtgtgaaaaactttagttcacttaaataacatttttgttgtttgtgttgtaaacagCGAAACGATCCAGACG
TTTTCAGAGAAgttatcagaattggatataatggtgagaagaatggtaatggaaagcttcgggatagagaagtaccttg
acaaacacctgaactcaacgaattaccgtctgcggctgatgaagtatatagcaccgcctgatgctgatgctactaatgt
tgcggctgatgccaaagatgctgatgataatgctaagacgattacaaatgataaagttgatgcggctggtgctaatgat
gtagatgctggtgatatcgctaatggtattgctaatcttcatattggtgatgatgctaacgctggtgctaatggtgctg
gtgttgatgctaatgatggtggtgaggatgctaagactggtgaggatgctaagactggtgaatgtgctagtgttaagtc
taatgccgaagatggtactgatgttaatgccagTGCTGATGCTGGTGTTACTGTTGGCTCTAATGCTGATGCTAATGCT
AATGCTAATGCTAATACTAGTACTGATGCTGGTGTTGGCGATAGTGTTAAAGCTAATGGTGGTGCTGATGATGTTGAGA
AGAAATTGGGTCTACCTTCTCACACTGATAAGAACCTTATAACGGTGCTTTATCAATACGAGATTGAAGGCTTGGAGGT
TCTAACCAAAGATGACAAGTGGATCAGACTCAAACCATCTCATAATTCTTTCGTTGTTATGGCTGGAGATTCTCTATAC
gtaagtttccaacttcttcttcttcttttttcttttttttaagttgacactcacacgtactgacgtacacgttggtgg
atttaaaagtaaccctagtggagaagaagatgaattttcatttacattatatcataaacctacttttaaattagaata
agaataattaaaactaaacccatttttattggctcactatggcctaaagaatataattaaaatattatataggctcaa
taagtcataatattctttagcctatagaatatttttaaagtattcaataattaaaattattttatagcatatagaatat
tttatgggctcaataattacaagtattctacaggatcaccatggcctaaagaataatcaaaagtaaaccgaatttaaa
attacaggtatagagaaagaaaaaagaaagctaaaattaaaaccaagaataagttaaaaatgtatgagaagtaacaaac
ttagtggcgaaaaagagaaaaagattattattcagtcacgttcacgctcactatggaccaacttatcctatagaaacta
ttaatattttcttgattttattcgctcttatcactttcacaagtgcatgtttgactaaaagcgttataacatgatgttt
tgttttcttctgatcatgatctcttgctgttacttacaaacaaaacaaatggtgattttgtttttgtttttttttgcag
GCACTTATGAATGGTAGACTAACTCGTCCCTTTCATCGAGTAAGAGTAACGGAGAAAAAGAAGACAAGATATTCAATAG

CATTGTTCTCGGCTCCAACCGCAGATTACATCATAGACACACCAAAAGAACTTGTGGACGAGAAGCATCCACGTATCTT
CGAACCATTTAACTATAACGACTTGATGAGTTTCTATCATAGTGAAGCTGGTCGTAAAGCTCGATCTACTCTTGATGCT
TTCTGTGCCGTCTCTCGAGCATAA

Ta FAE1

Ta-fae1-1 **(highlighted red; substitution: C to T; cDNA pos:1018)**

Ta-fae1-2 **(highlighted red; substitution: G to A; cDNA pos:1349)**

ATGACGTCCGTTAACGTTAAGCTCCTTTACCATTACGTCATCACCAACTTTTTCAACCTTTGCTTCTTCCCGTTAGCGG
CGATCGTTGCCGGAAAAGCCTCTCGGCTTACCACAAACGATCTTCACCACTTCTACTATTCCTATCTCCAACACAACCT
AATAACCATATCTCTACTCTTTGCCTTCACCGTTTTCGGTTTGGCTCTCTACATCGTAACCCGGCCCAAACCGGTTTAC
CTCGTTGACCATTCCTGCTACCTTCCACCATCGCATCTTAGAAGCAGTATCTCTAAGGTCATGGATATCTTCTATCAAG
TAAGATTAGCCGATCCTTTACGGAACGCGGCAAGCGATGATTCGTCCTGGCTTGATTTCTTGAGGAAGATTCAGGAGCG
GTCTGGTCTAGGCGATGAAACCCACGGCCCCGAGGGACTGCTTCAGGTCCCTCCACGGAAGACTTTTGCCGCGGCGCGT
GAAGAAACAGAGCAAGTGATCATCGGTGCGCTCGAAAAACTATTCGAGAACACCAAAGTTAACCCTAAAGAGATTGGTA
TACTTGTGGTGAACTCAAGCATGTTTAATCCGACTCCTTCGCTCTCGGCGATGGTTGTTAATACTTTCAAGCTCCGAAG
CAACATCAGAAGCTTTAATCTTGGAGGAATGGGTTGTAGTGCCGGCGTTATAGCCATTGATCTGGCTAAGGACTTGTTG
CATGTCCATAAAAACACTTATGCTCTTGTGGTGAGCACAGAGAACATCACTTACAACATTTATGCTGGTGATAACAGAT
CCATGATGGTTTCGAATTGCTTGTTCCGTGTTGGTGGGGCCGCGATTTTGCTCTCCAACAAGCCGAGGGACCGGAGACG
GTCCAAGTACCAGCTACTTCACACGGTTCGGACGCATACCGGAGCTGACGACAAGTCTTTCCGATGTGTGCAACAAGAA
GACGACGAGAGCGGTAAAACCGGGGTGTGTTTGTCCAAGGACATAACCGGTGTTGCCGGGAGAACTGTTCAGAAAAACA
TAACAACATTGGGTCCGTTGGTTCTTCCTTTTAGCGAGAAATTTCTTTTTTTCGTTACCTTCATCGCCAAGAAACTCTT
TAAAGACAAGATCAAACATTACTACGTCCCGGATTTCAAGCTTGCTATCGACCATTTTTGTATTCATGCCGGAGGCAGA
GCCGTGATCGATGTGCTACAGAAGAACTTAGGTCTATTGCCGATCGATGTGGAGGCATCTAGGTCAACGTTACATAGAT
TTGGGAACACTTCGTCTAGCTCAATTTGGTATGAATTGGCGTACATAGAGGCAAAAGGAAGGATGAAGAGAGGGAACAA
AGTTTGGCAGATTGCTTTAGGGTCAGGGTTTAAGTGTAATAGTGCGGTTTGGGTGGCTCTACGCAATGTCAAGGCTTCG
ACAAATAGTCCTTGGGAACATTGCATTGATAGATATCCAGATGCAATTGATTCTGATTCGGGTAAGTCAGAGACTCGTG
TCCAAAACGGTCGGTCCTAA

Ta ROD1

Ta-rod1-1 **(highlighted red; substitution: C to T; cDNA pos:678)**

Ta-rod1-2 **(highlighted red; substitution: C to T; cDNA pos:743)**

ATGTCAACTAAAACCGTCGTCCCTCTCCGTCGCAGATCTAAGCCCCTTAACGGAAATCACACTAACGGCGTCGCCATTG
ACGGAAGCCTCGACGACGACCACAACCGTCGCATCGGATCAGTAAATAGCCAAATGGATAACATTGCTAAGAAAACGGA
CGACGGCTACGCAAACGGCGGAGGAGGAGGAGGAGGAGGGAAAAGCAAGGCGTCGTTTATGACGTGGACGGCGCGTGAC
GTTGTGTACGTGGCGAGGTACCATTGGATACCGTGTTTGTTCGCGGTCGGGGTTCTGTTCTTCACGGGCGTGGAGTACA
CGCTCCAGATGATTCCCGCGAGGTCTGAGCCGTTCGATATTGGGTTTGTGGCCACGCGCTCTCTGAATCGCGTCTTGGC
AAATTCACCGGATCTTAACACCGTCTTAGCCGCTCTAAACACGgtaatttcgtactaattaatttagggtaaaaaatat
agtatttaataatgactatcctcaattccttcatgcttcacctaatattttgttttttttcgttgtcattaaaatcgt
aataatatattgagttagtcaaatgaaaaaaacaagtggcggtagtgattggaaacaaatctcagatcttttatctgtt
taataaggtatttaattatccagctggaattatgctgtcaagtgtcaacacagtagtagtaacatgcaatggaatttct
caatagaaaaaggtcttaattagtatagataattagtggacaaaaatgtagttaatgtaatctctttgctaagtagtta
tcataatcatctttttaacaactgccattttgtctgtgtgtttgttttacaacgaagtagtagtagaatagatcgcttt
ttagcttttgaaagtttcgaacccaaggaaaagggacacatgggttatgagttggagacacgatcacatgcaaacagag
agattggttaaattatcgacttttttgtagtacttttttaaaaaaaaactatttatataaaaaacatggtggatggtgggg
acagGTGTTCGTAGGGATGCAAACGACGTATATTGTATGGACATGGTTAATGGAAGGACGACCACGAGCCACCATCTCG
GCTTGCTTCATGTTTACTTGTCGAGGCATTCTTGGTTACTCTACTCAGCTCCCTCTTCCTCAGgttccaatcaacactt
ttcttctatctcttttcttaattaaaataattaccaattaactaaatgctaatcagtcgatatatcatagttccaacgt
tttggacgtgtgatttccattggccactaccatataaaacaacagagtctctttattcattattcaatatatatttgag
tattgatattattcatagggaggtttcatttgtactatcaataaaatttctacaactcttggatttttttctgctacatt
ttgtagttatttttttaattacttttaaaaacttgtgaataggagagactaatagtagtacgtaatatgattgtatcaa
atgctttaacatgtggggtttgggttaactatcatcatttcatagatcactattttgttttcgtttgttacctaacttt
ttgttatctttgaaaaataatgttccacgagttgattgactggacataaaaatcagattctctcactcatttacgttct
acggttctagccactcgttttttttctttttctttctgtggtgtaacacgtagataatggattttctatgtgtgtcgtct
tgctcaagaataataaatgtggttaaaggttaaatatagctctggaaattaattatctcctctttttttattaaccagG
ATTTTCTAGGATCAGGTGTCGATTTTCCGGTGGGAAACGTCTCGTTCTTCCTCTTCTACTCGGGTCACGTCGCCGGTTC
GATGATCGCATCTTTGGACATGAGGAGAATGCAGAGGATGAGACTAGCGATGCTTTTTGACATCCTCAATGTATTACAA

1100 TCGATCAGGCTGCTCGGGACGAGAGGACACTACACGATTGATCTCGCTGTCGGAGTTGGCGCTGGGATTCTCTTTGATT
1101 CATTCGCCGGCAAGTACGAAGAGATGATAAGCAAGAGACACAATTTAGTCAATGGTTTTGGTTTGATTTCGAAAGACTC
1102 GCTAGTCAATTAA
1103
1104

Data S2. Predicted protein sequences of the candidate genes for pennycress domestication obtained from Thlaspi version 1.0 gene annotations and mutation and information on the mutation sites.

Ta ELF6

Ta-elf6-1 **(bold red; Substitution: His to Tyr; AA pos:318)**

MGDVEIPNWLKALPLAPVFRPTDTEFADPIAYISKIEKEASAFGICKIIPPLPKPSKKYVFYNLNKSLLRCPELASDVD
ISKVCQEDRAVFTTRQQELGQAVKRKKGGESSKSNSQRSGVKQVWQSGGVYTLEQFESKSKTFYKSQLGTTKEVPPVVV
EALFWKAALEKPIYIEYANDVPGSAFGEPEGHFRHFRQRKRRGRGFYQRKAEVSEDSGVENGTNSQEPTCKNGEKTLPE
VAKASLASPSLLSQDPSKQKNMDIVDEMEGTAGWKLSNSSWNLQMIARSPGSVTRFMPDDIPGVTSPMVYIGMLFSWFA
W**H**VEDHELHSMNYLHTGSPKTWYAVPGDYAFDFEEVIRKNSYGRNIDQLAALTQLGEKTTLVSPEMIIASDIPCCRLVQ
NPGEFVVTFPRSYHVGFSHGFNCGEAANFGTPQWLNVAKEAAVRRAAMNYLPMLSHQQLLYLLTMSFVSRQISMASLVP
RSLLPGGRSSRLRDRQREEREFLVKKAFVEDILNENKNLSVLHREPGFRLVMWDPDLLPRHSVHGLVTVGGAAVSSPAE
GKNELEEKNKEKTTLLEELSLFMEKLKDVYYDDDDGLLNDFQVDSGTLACVACGVLGFPFMSVVQPSENALNDLSERRG
EIDGQEITALLSEKSDCEWNMSSRYIRPRIFCLEHTIELQRLLESRGGLKFLVICHKDFQKFKAYAAIVAEEVKVPFSY
DDILLESASKEELSLIDLAIEDEENNEHGVDWTSKLGINLRYCVKVRKNSPSTKIQHALSLGGLFSDTNHMLDMSTIKW
LQRKSRSKAKPSCTSSFTPREHLEVKVDRKLGEKEKVESQAGRKEEKIIQYSRKKKLKPKPSEERSQELTISAKSEDFE
NTCNTLAKRSHHHGAMHSDMNNEIGDFGRNGVSFSENHCSSPFTGARGQEHPKIIIKFGSALHGNITSSSSLVNGISAD
LTSVTREHQGHSMTSNNNGSNSSNHDGPIKLSGEHVSDVSVRDVDEAVEMSDQEFEELRSTVTNIEEEQQSEMVRPTAL
QVEGEESMCTREILSSEDIMHTEQQQEQTQLGLEVPETDIASENIVVDMIHDDEPLATRDILSSSNGDQASSNGLQALD
NELSMESEVASSENTEVIEASPNSIMREANKKRRIESESETNDNPDGSIGFIRSPCEGLRSRGRRRVTREASVSLTETS
DEEKKPAAKRFKKTPKTRSGSHHQEDSTTSHHNRCNLEGCKMTFKSKAELQAHQRNRCAHEGCGKKFRAHKYLVLHQRV
HNDDRPFVCSWKGCSMTFKWPWARTEHLRLHTGERPYKCKVDGCGMSFRFVSDYSRHRRKKGHYVT

Ta IND

Ta-ind-1 **(bold red; Substitution: Pro to Ser; AA pos:83)**

Ta-ind-2 **(bold red; Substitution: Arg to His; AA pos:87)**

Ta-ind-3 **(bold red; Substitution: Ala to Thr; AA pos:101)**

MNWNKPNDLITQEYPFLHDPHLMIDPPPETLSHFQPPPTLFSGHGGEEEEEEDNEEEEMDAMKEMQYTIAAMQPVDIDP
ATV**P**KPN**R**RNVRVSDDTQTVV**A**RRRREKISEKIRILKRMVPGGAKMDTASMLDEAIRYTKFLKRQVKLLQPHSQLGAPM
SDPSCLCYYHNSQT*

Conserved bHLH region is underlined.

Ta AOP2-like

Ta-aop2-1 **(bold red; Substitution: Gly to Arg; AA pos:97)**

MGSLSNTPQLPVIYLSDQTLKPGSSKWVEVRSDVRKALEEYGGFEVSYDRVSEELKKSVLQAMEELFALPVEAKQRNVS
PKPFSGYSTHNGLSESM**G**IQDPHVLDKVYEFTQLLRPDHCDGNKSISETIQTFSENADAGVTVGSNADANANANANTST
DAGVGDSVKANGGADDVEKKLGLPSHTDKNLITVLYQYEIEGLEVLTKDDKWIRLKPSHNSFVVMAGDSLYALMNGRLT
RPFHRVRVTEKKKTRYSIALFSAPTADYIIDTPKELVDEKHPRIFEPFNYNDLMSFYHSEAGRKARSTLDAFCAVSRA

Ta FAE1

Ta-fae1-1 **(bold red; Substitution: Gln to stop; AA pos:340)**

Ta-fae1-2 **(bold red; Substitution: Trp to stop; AA pos:450)**

MTSVNVKLLYHYVITNFFNLCFFPLAAIVAGKASRLTTNDLHHFYYSYLQHNLITISLLFAFTVFGLALYIVTRPKPVY
LVDHSCYLPPSHLRSSISKVMDIFYQVRLADPLRNAASDDSSWLDFLRKIQERSGLGDETHGPEGLLQVPPRKTFAAAR
EETEQVIIGALEKLFENTKVNPKEIGILVVNSSMFNPTPSLSAMVVNTFKLRSNIRSFNLGGMGCSAGVIAIDLAKDLL
HVHKNTYALVVSTENITYNIYAGDNRSMMVSNCLFRVGGAAILLSNKPRDRRRSKYQLLHTVRTHTGADDKSFRCVQQE
DDESGKTGVCLSKDITGVAGRTV**Q**KNITTLGPLVLPFSEKFLFFVTFIAKKLFKDKIKHYYVPDFKLAIDHFCIHAGGR
AVIDVLQKNLGLLPIDVEASRSTLHRFGNTSSSSIWYELAYIEAKGRMKRGNKV**W**QIALGSGFKCNSAVWVALRNVKAS
TNSPWEHCIDRYPDAIDSDSGKSETRVQNGRS*

Ta ROD1

Ta-rod1-1 **(bold red; Substitution: Met to Ile; AA pos:226)**

Ta-rod1-2 **(bold red; Substitution: Arg to Lys; AA pos:248)**

MSTKTVVPLRRRSKPLNGNHTNGVAIDGSLDDDHNRRIGSVNSQMDNIAKKTDDGYANGGGGGGGGKSKASFMTWTARD
VVYVARYHWIPCLFAVGVLFFTGVEYTLQMIPARSEPFDIGFVATRSLNRVLANSPDLNTVLAALNTVFVGMQTTYIVW
TWLMEGRPRATISACFMFTCRGILGYSTQLPLPQDFLGSGVDFPVGNVSFFLFYSGHVAGSMIASLD**M**RRMQRMRLAML
FDILNVLQSI**R**LLGTRGHYTIDLAVGVGAGILFDSFAGKYEEMISKRHNLVNGFGLISKDSLVN*

Data S3. Protein sequence alignments of *Thlaspi arvense* (pennycress) wild-type sequences with the corresponding mutant and orthologous Arabidopsis sequences. At the locations of the mutations, wild-type amino acids are highlighted in red and mutant amino acids in blue.

ELF6 - EARLY FLOWERING 6 (AT5G04240)

```
At-ELF6         MGNVEIPNWLKALPLAPVFRPTDTEFADPIAYISKIEKEASAFGICKIIPPLPKPSKKYV
Ta-ELF6         MGDVEIPNWLKALPLAPVFRPTDTEFADPIAYISKIEKEASAFGICKIIPPLPKPSKKYV
Ta-elf6-1       MGDVEIPNWLKALPLAPVFRPTDTEFADPIAYISKIEKEASAFGICKIIPPLPKPSKKYV
                **:************************************************************

At-ELF6         FYNLNKSLLKCPELVSDVDISKVCKEDRAVFTTRQQELGQTVKKNK-GEKGKSNSQRSGV
Ta-ELF6         FYNLNKSLLRCPELASDVDISKVCQEDRAVFTTRQQELGQAVKRKKGGESSKSNSQRSGV
Ta-elf6-1       FYNLNKSLLRCPELASDVDISKVCQEDRAVFTTRQQELGQAVKRKKGGESSKSNSQRSGV
                *********:****.*********:***************:**::* **..*********

At-ELF6         KQVWQSGGVYTLDQFEAKSKAFYKTQLGTVKELAPVVIEALFWKAALEKPIYIEYANDVP
Ta-ELF6         KQVWQSGGVYTLEQFESKSKTFYKSQLGTTKEVPPVVVEALFWKAALEKPIYIEYANDVP
Ta-elf6-1       KQVWQSGGVYTLEQFESKSKTFYKSQLGTTKEVPPVVVEALFWKAALEKPIYIEYANDVP
                ************:***:***:***:****.**: ***:**********************

At-ELF6         GSAFGEPEDHFRHFRQRKRRGRGFYQRKTENN------------DPSGKNGEKSSPEVEK
Ta-ELF6         GSAFGEPEGHFRHFRQRKRRGRGFYQRKAEVSEDSGVENGTNSQEPTCKNGEKTLPEVAK
Ta-elf6-1       GSAFGEPEGHFRHFRQRKRRGRGFYQRKAEVSEDSGVENGTNSQEPTCKNGEKTLPEVAK
                ********.******************:* .            :*: *****: *** *

At-ELF6         APLASTSLSSQDSSKQKNMDIVDEMEGTAGWKLSNSSWNLQMIARSPGSVTRFMPDDIPG
Ta-ELF6         ASLASPSLLSQDPSKQKNMDIVDEMEGTAGWKLSNSSWNLQMIARSPGSVTRFMPDDIPG
Ta-elf6-1       ASLASPSLLSQDPSKQKNMDIVDEMEGTAGWKLSNSSWNLQMIARSPGSVTRFMPDDIPG
                * *** ** *** ************************************************

At-ELF6         VTSPMVYIGMLFSWFAWHVEDHELHSMNYLHTGSPKTWYAVPCDYALDFEEVIRKNSYGR
Ta-ELF6         VTSPMVYIGMLFSWFAWHVEDHELHSMNYLHTGSPKTWYAVPGDYAFDFEEVIRKNSYGR
Ta-elf6-1       VTSPMVYIGMLFSWFAWYVEDHELHSMNYLHTGSPKTWYAVPGDYAFDFEEVIRKNSYGR
                *****************:************************ ***:**************

At-ELF6         NIDQLAALTQLGEKTTLVSPEMIVASGIPCCRLVQNPGEFVVTFPRSYHVGFSHGFNCGE
Ta-ELF6         NIDQLAALTQLGEKTTLVSPEMIIASDIPCCRLVQNPGEFVVTFPRSYHVGFSHGFNCGE
Ta-elf6-1       NIDQLAALTQLGEKTTLVSPEMIIASDIPCCRLVQNPGEFVVTFPRSYHVGFSHGFNCGE
                ***********************:**.**********************************

At-ELF6         AANFGTPQWLNVAKEAAVRRAAMNYLPMLSHQQLLYLLTMSFVSR-------VPRSLLPG
Ta-ELF6         AANFGTPQWLNVAKEAAVRRAAMNYLPMLSHQQLLYLLTMSFVSRQISMASLVPRSLLPG
Ta-elf6-1       AANFGTPQWLNVAKEAAVRRAAMNYLPMLSHQQLLYLLTMSFVSRQISMASLVPRSLLPG
                *******************************************       ********

At-ELF6         GRSSRLRDRQREEREFLVKRAFVEDILNENKNLSVLLREPGSRLVMWDPDLLPRHSALAL
Ta-ELF6         GRSSRLRDRQREEREFLVKKAFVEDILNENKNLSVLHREPGFRLVMWDPDLLPRHSVHGL
Ta-elf6-1       GRSSRLRDRQREEREFLVKKAFVEDILNENKNLSVLHREPGFRLVMWDPDLLPRHSVHGL
                ******************:**************** **** **************. .*

At-ELF6         AAAGVAGASAVSPPAVAKKELEEGHSELQNKEKTSLLEELSLFMEKLNDVYYDDDDGLLN
Ta-ELF6         VTVG---GAAVSSPAEGKNELE-----EKNKEKTTLLEELSLFMEKLKDVYYDDDDGLLN
Ta-elf6-1       VTVG---GAAVSSPAEGKNELE-----EKNKEKTTLLEELSLFMEKLKDVYYDDDDGLLN
                .:.*   .:*** ** .*:***      :*****:************:*************

At-ELF6         DFQVDTGTLPCVACGVLGFPFMSVVQPSEKALKDLSERQGETDAQEIMTLSSEKSDCEWK
Ta-ELF6         DFQVDSGTLACVACGVLGFPFMSVVQPSENALNDLSERRGEIDGQEITALLSEKSDCEWN
Ta-elf6-1       DFQVDSGTLACVACGVLGFPFMSVVQPSENALNDLSERRGEIDGQEITALLSEKSDCEWN
                *****:*** ******************:**:*****:** *.*** :* ********:
```

```
At-ELF6        TSSRYIRPRIFCLEHTIELQRLLQSRGGLKFLVICHKDFQKFKAHAAIVAEEVKVPFSYD
Ta-ELF6        MSSRYIRPRIFCLEHTIELQRLLESRGGLKFLVICHKDFQKFKAYAAIVAEEVKVPFSYD
Ta-elf6-1      MSSRYIRPRIFCLEHTIELQRLLESRGGLKFLVICHKDFQKFKAYAAIVAEEVKVPFSYD
                ***********************:*********************:***************

At-ELF6        DVLLESASQEELSLIDLAIEDEEKYEHSVDWTSELGINLRYCVKVRKNSPTKKIQHALSL
Ta-ELF6        DILLESASKEELSLIDLAIEDEENNEHGVDWTSKLGINLRYCVKVRKNSPSTKIQHALSL
Ta-elf6-1      DILLESASKEELSLIDLAIEDEENNEHGVDWTSKLGINLRYCVKVRKNSPSTKIQHALSL
               *:******:***************: **.*****:****************:.********

At-ELF6        GGLFSDTSQMLDFTTIRWLQRKSRSKAKPSSTSSFTPCEHLEVKADGKLR--DNLDSQTG
Ta-ELF6        GGLFSDTNHMLDMSTIKWLQRKSRSKAKPSCTSSFTPREHLEVKVDRKLGEKEKVESQAG
Ta-elf6-1      GGLFSDTNHMLDMSTIKWLQRKSRSKAKPSCTSSFTPREHLEVKVDRKLGEKEKVESQAG
               *******.:***::**:*************.****** ******.* **   ::::**:*

At-ELF6        KKEEKIIQYSRKKKLNPKPSAEQVQELATLAKSKDFDKTCKNFSSRSHLDSAIRSEMNSE
Ta-ELF6        RKEEKIIQYSRKKKLKPKPSEERSQELTISAKSEDFENTCNTLAKRSHHHGAMHSDMNNE
Ta-elf6-1      RKEEKIIQYSRKKKLKPKPSEERSQELTISAKSEDFENTCNTLAKRSHHHGAMHSDMNNE
               :***************:**** *: ***:  ***:**::**:.::.*** ..*::*:**.*

At-ELF6        IGDSGRVIGVSFSINPCSSSFTVGHGQEHPEITVKFGSDLDGNVTNSLSMVNGDSADLTL
Ta-ELF6        IGDFGR-NGVSFSENHCSSPFTGARGQEHPKIIIKFGSALHGNITSSSSLVNGISADLTS
Ta-elf6-1      IGDFGR-NGVSFSENHCSSPFTGARGQEHPKIIIKFGSALHGNITSSSSLVNGISADLTS
               *** **  ***** * *** ** .:*****:* :**** *.**:*.* *:*** *****

At-ELF6        TSISREQHQGHSMTSNNNGSNSGSHVVASQTILVSTGDNHDGPRKLSGDYVCSDVSVRGI
Ta-ELF6        V---TREHQGHSMTSNNNGSNS---------------SNHDGPIKLSGEHV-SDVSVRDV
Ta-elf6-1      V---TREHQGHSMTSNNNGSNS---------------SNHDGPIKLSGEHV-SDVSVRDV
               .    .:***************               .***** ****::* ******.:

At-ELF6        QEAVEMSDQEFGEPRSTVTNIEDEQQSQIVKPTQREAVFGDHEQVEGAEAVSTRENLCSE
Ta-ELF6        DEAVEMSDQEFEELRSTVTNIEEEQQSEMVRPTAL--------QVEGEESMCTREILSSE
Ta-elf6-1      DEAVEMSDQEFEELRSTVTNIEEEQQSEMVRPTAL--------QVEGEESMCTREILSSE
               :********** * ********:****::*:**          **** *:::*** *.**

At-ELF6        IILHTEHS--SAHVGMEIPDINTASENLVVDMTHDGEPLESSDILSSSNGDEASSNGLQV
Ta-ELF6        DIMHTEQQQEQTQLGLEVPETDIASENIVVDMIHDDEPLATRDILSSSNGDQASSNGLQA
Ta-elf6-1      DIMHTEQQQEQTQLGLEVPETDIASENIVVDMIHDDEPLATRDILSSSNGDQASSNGLQA
                *:***:.  .:::*:*:*: : ****:**** **.*** : *********:*******.

At-ELF6        LNDELSMESEVSSSENTEVIEAP--NSMGEAKKKRKIESESETNDNPESSIGFIRSPCEG
Ta-ELF6        LDNELSMESEVASSENTEVIEASPNSIMREANKKRRIESESETNDNPDGSIGFIRSPCEG
Ta-elf6-1      LDNELSMESEVASSENTEVIEASPNSIMREANKKRRIESESETNDNPDGSIGFIRSPCEG
               *::********:**********   . * **:***:***********:.***********

At-ELF6        LRSRGKRKATCETSLKHTETSDEEKKPIAKRLKKTPKACSGSRQQEVPTTTHPNRCYLEG
Ta-ELF6        LRSRGRRRVTREASVSLTETSDEEKKPAAKRFKKTPKTRSGSHHQEDSTTSHHNRCNLEG
Ta-elf6-1      LRSRGRRRVTREASVSLTETSDEEKKPAAKRFKKTPKTRSGSHHQEDSTTSHHNRCNLEG
               *****:*:.* *:*:. ********** ***:*****: ***::**  **:* *** ***

At-ELF6        CKMTFESKAKLQTHKRNRCTHEGCGKKFRAHKYLVLHQRVHKDERPFECSWKGCSMTFKW
Ta-ELF6        CKMTFKSKAELQAHQRNRCAHEGCGKKFRAHKYLVLHQRVHNDDRPFVCSWKGCSMTFKW
Ta-elf6-1      CKMTFKSKAELQAHQRNRCAHEGCGKKFRAHKYLVLHQRVHNDDRPFVCSWKGCSMTFKW
               *****:***:**:*:****:**********************:*:*** ************

At-ELF6        QWARTEHLRLHTGERPYICKVDGCGLSFRFVSDYSRHRRKTMHYVT
Ta-ELF6        PWARTEHLRLHTGERPYKCKVDGCGMSFRFVSDYSRHRRKKGHYVT
Ta-elf6-1      PWARTEHLRLHTGERPYKCKVDGCGMSFRFVSDYSRHRRKKGHYVT
                **************** *******:**************. ****
```

The JmjC domain is highlighted with bold lettering.

IND – INDEHISCENT (AT4G00120)

```
At-IND        MENGMYKKKGVCDSCVSSKSRSNHSPKRSMMEPQPHHLLMDWNKANDLLTQEHAAFLNDP
Ta-IND        ------------------------------------MNWNKPNDLITQEY-PFLHDP
Ta-ind-1      ------------------------------------MNWNKPNDLITQEY-PFLHDP
Ta-ind-2      ------------------------------------MNWNKPNDLITQEY-PFLHDP
Ta-ind-3      ------------------------------------MNWNKPNDLITQEY-PFLHDP
                                                   *:*** ***:***:  **:**

At-IND        HHLMLDPPPETLIHLDE--------------DEEYDEDMDAMKEMQYMIAVMQPVDIDP
Ta-IND        -HLMIDPPPETLSHFQPPPTLFSGHGGEEEEEEDNEEEEMDAMKEMQYTIAAMQPVDIDP
Ta-ind-1      -HLMIDPPPETLSHFQPPPTLFSGHGGEEEEEEDNEEEEMDAMKEMQYTIAAMQPVDIDP
Ta-ind-2      -HLMIDPPPETLSHFQPPPTLFSGHGGEEEEEEDNEEEEMDAMKEMQYTIAAMQPVDIDP
Ta-ind-3      -HLMIDPPPETLSHFQPPPTLFSGHGGEEEEEEDNEEEEMDAMKEMQYTIAAMQPVDIDP
               ***:******* *::               ::: :*:********* **.********

At-IND        ATVPKPNRRNVRISDDPQTVVARRRRERISEKIRILKRIVPGGAKMDTASMLDEAIRYTK
Ta-IND        ATVPKPNRRNVRVSDDTQTVVARRRREKISEKIRILKRMVPGGAKMDTASMLDEAIRYTK
Ta-ind-1      ATVSKPNRRNVRVSDDTQTVVARRRREKISEKIRILKRMVPGGAKMDTASMLDEAIRYTK
Ta-ind-2      ATVPKPNKRNVRVSDDTQTVVARRRREKISEKIRILKRMVPGGAKMDTASMLDEAIRYTK
Ta-ind-3      ATVPKPNRRNVRVSDDTQTVVTRRRREKISEKIRILKRMVPGGAKMDTASMLDEAIRYTK
              *** ***:****:*** ****:*****:**********:**********************

At-IND        FLKRQVRILQPHSQIGAPMANPSYLCYYHNSQP
Ta-IND        FLKRQVKLLQPHSQLGAPMSDPSCLCYYHNSQT
Ta-ind-1      FLKRQVKLLQPHSQLGAPMSDPSCLCYYHNSQT
Ta-ind-2      FLKRQVKLLQPHSQLGAPMSDPSCLCYYHNSQT
Ta-ind-3      FLKRQVKLLQPHSQLGAPMSDPSCLCYYHNSQT
              ******::******:****::** ********
```

AOP2 – alkenyl- and hydroxyalkyl-producing (AT4G03070)

```
At-AOP2        MDSDFVPPSVSFQLPVIDFSDQNLKPGSSKWDEVTADVLKALEDYGCFEASFDKLSVELN
Ta-AOP2        ----MGSLSNTPQLPVIYLSDQTLKPGSSKWVEVRSDVRKALEEYGGFEVSYDRVSEELK
Ta-aop2        ----MGSLSNTPQLPVIYLSDQTLKPGSSKWVEVRSDVRKALEEYGGFEVSYDRVSEELK
                   :   * : ***** :***.******** ** :** ****:** **.*:*::* **:

At-AOP2        RSVFEAMEDLFELPIPTKQRNVSSKPFHGYLCHN-LYESLGINDANVLEKVNDFTQQLWP
Ta-AOP2        KSVLQAMEELFALPVEAKQRNVSPKPFSGYSTHNGLSESMGIQDPHVLDKVYEFTQLLRP
Ta-aop2        KSVLQAMEELFALPVEAKQRNVSPKPFSGYSTHNGLSESMRIQDPHVLDKVYEFTQLLRP
               :**::***:** **: :****** *** **  ** * **: *:* :**:** :*** * *

At-AOP2        DH--GNKSISETIHLFSEQLVELDLMVRRMIMESFGIENYIDEHLNSTYY-LTRLMKYTS
Ta-AOP2        DHCDGNKSISETIQTFSENADAGVTV---------GS--NADANANANANTSTDAGVGDS
Ta-aop2        DHCDGNKSISETIQTFSENADAGVTV---------GS--NADANANANANTSTDAGVGDS
               **  *********: ***:      :         *     * : *:.    *      *

At-AOP2        PPDDDDDDDEETKLGLRSHTDKNIITILHQYQVDGLEVKTKDDKWIKVKPSQDSVLVMVG
Ta-AOP2        VKANGGADDVEKKLGLPSHTDKNLITVLYQYEIEGLEVLTKDDKWIRLKPSHNSFVVMAG
Ta-aop2        VKANGGADDVEKKLGLPSHTDKNLITVLYQYEIEGLEVLTKDDKWIRLKPSHNSFVVMAG
               :.. ** *.**** ******:**:*:**:::**** *******::***::*.:**.*

At-AOP2        DSLCALLNGRLHSPYHRVIMTG-KKTRYSTGLFSIPKTGVIIDSPEELVDKEHPRIFKPF
Ta-AOP2        DSLYALMNGRLTRPFHRVRVTEKKKTRYSIALFSAPTADYIIDTPKELVDEKHPRIFEPF
Ta-aop2        DSLYALMNGRLTRPFHRVRVTEKKKTRYSIALFSAPTADYIIDTPKELVDEKHPRIFEPF
               *** **:****  *:*** :*  ****** .*** *.:. ***:*:****::*****:**

At-AOP2        EYTDFLHFFQTEAGRIAQSALHAFAAF---
Ta-AOP2        NYNDLMSFYHSEAGRKARSTLDAFCAVSRA
Ta-aop2        NYNDLMSFYHSEAGRKARSTLDAFCAVSRA
```

```
                :*.*:: *:::**** *:*:*.**.*.
```

FAE1 – FATTY ACID ELONGATION1 (AT4G34520)

```
At-FAE1         MTSVNVKLLYRYVLTNFFNLCLFPLTAFLAGKASRLTINDLHNF-LSYLQHNLITVTLLF
Ta-FAE1         MTSVNVKLLYHYVITNFFNLCFFPLAAIVAGKASRLTTNDLHHFYYSYLQHNLITISLLF
Ta-fae1-1       MTSVNVKLLYHYVITNFFNLCFFPLAAIVAGKASRLTTNDLHHFYYSYLQHNLITISLLF
Ta-fae1-2       MTSVNVKLLYHYVITNFFNLCFFPLAAIVAGKASRLTTNDLHHFYYSYLQHNLITISLLF
                **********:**:*******:***:*::******** ****:*  *********::***

At-FAE1         AFTVFGLVLYIVTRPNPVYLVDYSCYLPPPHLKVSVSKVMDIFYQIRKADTSSRNVACDD
Ta-FAE1         AFTVFGLALYIVTRPKPVYLVDHSCYLPPSHLRSSISKVMDIFYQVRLADP-LRNAASDD
Ta-fae1-1       AFTVFGLALYIVTRPKPVYLVDHSCYLPPSHLRSSISKVMDIFYQVRLADP-LRNAASDD
Ta-fae1-2       AFTVFGLALYIVTRPKPVYLVDHSCYLPPSHLRSSISKVMDIFYQVRLADP-LRNAASDD
                *******.*******:******:****** **: *:*********:* **   **.*.**

At-FAE1         PSSLDFLRKIQERSGLGDETYSPEGLIHVPPRKTFAASREETEKVIIGALENLFENTKVN
Ta-FAE1         SSWLDFLRKIQERSGLGDETHGPEGLLQVPPRKTFAAAREETEQVIIGALEKLFENTKVN
Ta-fae1-1       SSWLDFLRKIQERSGLGDETHGPEGLLQVPPRKTFAAAREETEQVIIGALEKLFENTKVN
Ta-fae1-2       SSWLDFLRKIQERSGLGDETHGPEGLLQVPPRKTFAAAREETEQVIIGALEKLFENTKVN
                 * *****************:.****::*********:*****:*******:********

At-FAE1         PREIGILVVNSSMFNPTPSLSAMVVNTFKLRSNIKSFNLGGMGCSAGVIAIDLAKDLLHV
Ta-FAE1         PKEIGILVVNSSMFNPTPSLSAMVVNTFKLRSNIRSFNLGGMGCSAGVIAIDLAKDLLHV
Ta-fae1-1       PKEIGILVVNSSMFNPTPSLSAMVVNTFKLRSNIRSFNLGGMGCSAGVIAIDLAKDLLHV
Ta-fae1-2       PKEIGILVVNSSMFNPTPSLSAMVVNTFKLRSNIRSFNLGGMGCSAGVIAIDLAKDLLHV
                *:*********************************:************************

At-FAE1         HKNTYALVVSTENITQGIYAGENRSMMVSNCLFRVGGAAILLSNKSGDRRRSKYKLVHTV
Ta-FAE1         HKNTYALVVSTENITYNIYAGDNRSMMVSNCLFRVGGAAILLSNKPRDRRRSKYQLLHTV
Ta-fae1-1       HKNTYALVVSTENITYNIYAGDNRSMMVSNCLFRVGGAAILLSNKPRDRRRSKYQLLHTV
Ta-fae1-2       HKNTYALVVSTENITYNIYAGDNRSMMVSNCLFRVGGAAILLSNKPRDRRRSKYQLLHTV
                *************** .****:*********************  *******:*:***

At-FAE1         RTHTGADDKSFRCVQQEDDESGKIGVCLSKDITNVAGTTLTKNIATLGPLILPLSEKFLF
Ta-FAE1         RTHTGADDKSFRCVQQEDDESGKTGVCLSKDITGVAGRTVQKNITTLGPLVLPFSEKFLF
Ta-fae1-1       RTHTGADDKSFRCVQQEDDESGKTGVCLSKDITGVAGRTV*------------------
Ta-fae1-2       RTHTGADDKSFRCVQQEDDESGKTGVCLSKDITGVAGRTVQKNITTLGPLVLPFSEKFLF
                ********************** *********.*** *:

At-FAE1         FATFVAKKLLKDKIKHYYVPDFKLAVDHFCIHAGGRAVIDELEKNLGLSPIDVEASRSTL
Ta-FAE1         FVTFIAKKLFKDKIKHYYVPDFKLAIDHFCIHAGGRAVIDVLQKNLGLLPIDVEASRSTL
Ta-fae1-1       ------------------------------------------------------------
Ta-fae1-2       FVTFIAKKLFKDKIKHYYVPDFKLAIDHFCIHAGGRAVIDVLQKNLGLLPIDVEASRSTL

At-FAE1         HRFGNTSSSSIWYELAYIEAKGRMKKGNKAWQIALGSGFKCNSAVWVALRNVKASANSPW
Ta-FAE1         HRFGNTSSSSIWYELAYIEAKGRMKRGNKVWQIALGSGFKCNSAVWVALRNVKASTNSPW
Ta-fae1-1       ------------------------------------------------------------
Ta-fae1-2       HRFGNTSSSSIWYELAYIEAKGRMKRGNKV*----------------------------

At-FAE1         QHCIDRYPVKIDSDLSKSKTHVQNGRS-
Ta-FAE1         EHCIDRYPDAIDSDSGKSETRVQNGRS*
Ta-fae1-1       ----------------------------
Ta-fae1-2       ----------------------------
```

ROD1 – REDUCED OLEATE DESATURATION 1 (AT3G15820)

```
At-ROD1        MSAAAAETDVSLRRRSNSLNGNHTNGVAIDGTLDN-NNRRVGDTNTHMDISAKKTDNGYA
Ta-ROD1        ---MSTKTVVPLRRRSKPLNGNHTNGVAIDGSLDDDHNRRIGSVNSQMDNIAKKTDDGYA
Ta-rod1-1      ---MSTKTVVPLRRRSKPLNGNHTNGVAIDGSLDDDHNRRIGSVNSQMDNIAKKTDDGYA
Ta-rod1-2      ---MSTKTVVPLRRRSKPLNGNHTNGVAIDGSLDDDHNRRIGSVNSQMDNIAKKTDDGYA
                   :::* * *****: ************:**: :***:*..*::**  *****:***

At-ROD1        NGVGG-GGWRSKASFTTWTARDIVYVVRYHWIPCMFAAGLLFFMGVEYTLQMIPARSEPF
Ta-ROD1        NGGGGGGGGKSKASFMTWTARDVVYVARYHWIPCLFAVGVLFFTGVEYTLQMIPARSEPF
Ta-rod1-1      NGGGGGGGGKSKASFMTWTARDVVYVARYHWIPCLFAVGVLFFTGVEYTLQMIPARSEPF
Ta-rod1-2      NGGGGGGGGKSKASFMTWTARDVVYVARYHWIPCLFAVGVLFFTGVEYTLQMIPARSEPF
               ** ** ** :***** ******:***.*******:**.*:*** ***************

At-ROD1        DLGFVVTRSLNRVLASSPDLNTVLAALNTVFVGMQTTYIVWTWLVEGRARATIAALFMFT
Ta-ROD1        DIGFVATRSLNRVLANSPDLNTVLAALNTVFVGMQTTYIVWTWLMEGRPRATISACFMFT
Ta-rod1-1      DIGFVATRSLNRVLANSPDLNTVLAALNTVFVGMQTTYIVWTWLMEGRPRATISACFMFT
Ta-rod1-2      DIGFVATRSLNRVLANSPDLNTVLAALNTVFVGMQTTYIVWTWLMEGRPRATISACFMFT
               *:***.*********.****************************:*** ****:* ****

At-ROD1        CRGILGYSTQLPLPQDFLGSGVDFPVGNVSFFLFFSGHVAGSMIASLDMRRMQRLRLAMV
Ta-ROD1        CRGILGYSTQLPLPQDFLGSGVDFPVGNVSFFLFYSGHVAGSMIASLDMRRMQRMRLAML
Ta-rod1-1      CRGILGYSTQLPLPQDFLGSGVDFPVGNVSFFLFYSGHVAGSMIASLDIRRMQRMRLAML
Ta-rod1-2      CRGILGYSTQLPLPQDFLGSGVDFPVGNVSFFLFYSGHVAGSMIASLDMRRMQRMRLAML
               **********************************:*************:*****:****:

At-ROD1        FDILNVLQSIRLLGTRGHYTIDLAVGVGAGILFDSLAGKYEEMMSKRH-LGTGFSLISKD
Ta-ROD1        FDILNVLQSIRLLGTRGHYTIDLAVGVGAGILFDSFAGKYEEMISKRHNLVNGFGLISKD
Ta-rod1-1      FDILNVLQSIRLLGTRGHYTIDLAVGVGAGILFDSFAGKYEEMISKRHNLVNGFGLISKD
Ta-rod1-2      FDILNVLQSIKLLGTRGHYTIDLAVGVGAGILFDSFAGKYEEMISKRHNLVNGFGLISKD
               **********:************************:*******:**** * .**.*****

At-ROD1        SLVN-
Ta-ROD1        SLVN*
Ta-rod1-1      SLVN*
Ta-rod1-2      SLVN*
               ****
```