1 **Spring flowering habit in field pennycress (*Thlaspi arvense*) has arisen multiple**
2 **independent times**
3
4 **Kevin M. Dorn[1], Evan B. Johnson[1], Erin Daniels[1], Donald L. Wyse[2], M. David**
5 **Marks[1]**
6 [1]Department of Plant and Microbial Biology, University of Minnesota,
7 1445 Gortner Avenue, 250 Biological Sciences Center, Saint Paul, Minnesota 55108
8 [2]Department of Agronomy and Plant Genetics, University of Minnesota,
9 411 Borlaug Hall, 1991 Upper Buford Circle, Saint Paul, Minnesota 55108
10

11 Email addresses:
12 KMD: dorn@umn.edu
13 EBJ: john6551@umn.edu
14 ED: danie711@umn.edu
15 DLW: wysex001@umn.edu
16 MDM: marks004@umn.edu
17

18 Correspondence to:
19 M David Marks
20 University of Minnesota
21 Department of Plant Biology
22 250 Biological Sciences Center
23 1445 Gortner Avenue
24 St Paul, Minnesota 55108
25 Marks004@umn.edu
26

27 Word Count, Summary: 200
28 Word Count, Introduction: 1050
29 Word Count, Materials and Methods: 1159
30 Word Count, Results: 1630
31 Word Count, Discussion: 910
32 Word Count, Acknowledgements: 98
33 Total Word Count: 5065
34
35 2 Figures (both in colour)
36 6 Supporting Figures
37 3 Supporting Tables
38
39
40
41
42
43

Summary

44

45   - Pennycress (*Thlaspi arvense* L.) is currently being developed as a new cold-season

46   oilseed crop. Like many Brassicaceae, pennycress can exhibit either a winter or spring

47   annual phenotype. In Arabidopsis, mutations in negative regulators of flowering,

48   including FLOWERING LOCUS C (FLC) and FRIGIDA can cause the transition to a

49   spring annual habit. The genetics underlying the difference between spring and winter

50   annual pennycress are currently unknown.

51

52   - Using whole genome sequencing across wild spring annual pennycress accessions, co-

53   segregation analyses, and comparative genomics approaches, we identify new alleles of

54   TaFLC and explore their geographic distribution.

55

56   - We report that loss of function mutations in TaFLC confer the spring annual phenotype

57   in pennycress. We have also identified four natural alleles of TaFLC that confer a spring

58   annual growth habit. The two spring annual FLC alleles present in European accessions

59   were only identified in accessions collected in Montana, USA.

60

61   - In pennycress, the spring annual habit has arisen several independent times. Accessions

62   harboring the two European alleles were introduced to North America, likely after the

63   species became a widespread on the continent. These findings provide new information

64   on the natural history of the introduction and spread of the spring annual phenotype in

65   pennycress.

66

67

68

69

70

71

72   Key Words: *Thlaspi arvense*, pennycress, Flowering Locus C, spring annual, winter

73   annual, whole genome sequencing

74   **Introduction**

75   *Thlaspi arvense* L. (field pennycress, pennycress herein) is currently a target for

76   domestication as a new, cold-hardy winter oilseed crop that can fit within the

77   corn/soybean rotation in the Midwestern United States (Dorn *et al.*, 2013; Sedbrook,

78   2014; Dorn *et al.*, 2015). As a winter annual crop, pennycress can be planted in the late

79   summer to early autumn, either into standing corn or immediately after corn harvest. The

80   pennycress crop establishes a robust vegetative cover prior to winter, providing important

81   ecosystem services such as limiting soil erosion and nutrient runoff. In the spring,

82   pennycress will flower and set seed, yielding upwards of 1,600 pounds per acre of oil-

83   rich seed in time for planting a crop of soybean (Phippen & Phippen, 2012; Sedbrook,

84   2014). Pennycress seed is high in oils that can be readily converted to biodiesel or jet fuel,

85   with the remaining high-protein seed meal having potential as animal feed or feedstock

86   for industrial uses (Moser *et al.*, 2009; Selling *et al.*, 2011; Evangelista *et al.*, 2012; Fan

87   *et al.*, 2013; Hojilla-Evangelista *et al.*, 2013).

88

89   Pennycress is a member of the Thlaspideae, a tribe of the Brassicaceae (Beilstein *et al.*,

90   2008). The Brassicaceae family is divided into three lineages. Pennycress resides in

91   lineage II, along with the Brassica genus, including the economically important oilseed

92   crops *Brassica rapa* and *Brassica napus*, but not the model species *Arabidopsis thaliana*

93   nor *Capsella rubella*, which are members of lineage I (Franzke *et al.*, 2011). Pennycress

94   is native to Eurasia and is considered naturalized to most temperate to subarctic regions

95   throughout the northern hemisphere, including all of the United States except for Hawaii

96   and Alabama, all provinces of Canada, as well as being considered naturalized in the

97   southern hemisphere, including Australia, New Zealand, and Argentina (Warwick, 2002).

98   Pennycress exhibits numerous traits that make it an attractive winter rotation crop

99   (Sedbrook, 2014). Of particular importance in the upper Midwest is the existence of

100   pennycress lines that complete their life cycle rapidly enough to fit within the

101   corn/soybean rotation. Central to this rapid spring development of pennycress are its

102   underlying growth habits. In wild populations, there are both winter and spring annual

103   pennycress (Best & McIntyre, 1972; McIntyre & Best, 1978), similar to many other

104   Brassicaceae species such as *Arabidopsis thaliana* (Stinchcombe *et al.*, 2004; Shindo *et*

105   *al.*, 2005), *Brassica rapa* (Wu *et al.*, 2012), *Camelina sativa* (Crowley, 1999), and

106   *Brassica napus* (Tadege *et al.*, 2001).

107

108   Throughout the decades of research on basic developmental questions in Arabidopsis,

109   and the expanding translational research in other Brassica crops, the underlying

110   molecular mechanisms controlling flowering time have been identified in many of these

111   species (Simpson & Dean, 2002; Amasino, 2005; Jung & Muller, 2009; Kim *et al.*, 2009).

112   In wild accessions of Arabidopsis, only a handful of mutations are responsible for

113   variation of flowering time and vernalization requirement (Burn *et al.*, 1993; Clarke &

114   Dean, 1994; Johanson *et al.*, 2000). Most notably, allelic variation in two key negative

115   regulators, FLOWERING LOCUS C (FLC) and/or FRIGIDA (FRI), underlies the key

116   difference between spring and winter annual plants (Johanson *et al.*, 2000; Michaels &

117   Amasino, 2001; Michaels *et al.*, 2003; Michaels *et al.*, 2004; Shindo *et al.*, 2005; Jiang *et*

118   *al.*, 2009). FLC encodes a MADS box transcription factor (Michaels & Amasino, 1999),

119   and inhibits flowering prior to vernalization by repressing the expression of

120   FLOWERING LOCUS T (FT) (Searle *et al.*, 2006). In Arabidopsis, *flc* null mutations

121   eliminate the vernalization requirement and impart the spring annual, rapid-flowering

122   phenotype (Michaels & Amasino, 1999). Allelic variation within FRI also impacts

123   flowering time and the vernalization requirement. Similar to FLC, Arabidopsis accessions

124   harboring loss of function mutations in FRI flower rapidly without vernalization

125   (Johanson *et al.*, 2000; Shindo *et al.*, 2005).

126

127   The expression of FLC is positively regulated by FRI (Michaels & Amasino, 1999), thus,

128   mutations in either of these vernalization-responsive negative regulators can lead to a loss

129   of vernalization requirement and a spring annual growth habit (Michaels & Amasino,

130   1999; Michaels & Amasino, 2001). The vernalization signal provided by the cold of

131   winter removes the repression on the transition to flowering through the epigenetic

132   silencing of FLC. Specifically, vernalization increases histone 3 K27 trimethylation

133   (H3K27me3) at FLC chromatin, reducing transcriptional activity (Sung *et al.*, 2006;

134   Finnegan & Dennis, 2007; Greb *et al.*, 2007; Coustham *et al.*, 2012; Yang *et al.*, 2014).

135   The vernalization-induced silencing of FLC releases the repression on the transition to

136   flowering, which permits the transition to reproductive growth (Searle *et al.*, 2006).

137   Additionally, the roles of the COOLAIR long noncoding antisense transcripts transcribed

138   from the AtFLC locus have been revealed as a key regulatory component in the cold-

139   induced regulation of FLC (Swiezewski *et al.*, 2009; Csorba *et al.*, 2014; Marquardt *et al.*,

140   2014).

141

142   While this extensive understanding of the molecular genetic pathways controlling

143   flowering time and vernalization in Arabidopsis has informed similar studies in Brassica

144   relatives, little is known about the underlying mechanisms controlling flowering time

145   variation in pennycress. Different accessions of pennycress have been reported to either

146   act as early flowering or late flowering, with late flowering accessions growing as

147   rosettes for as long as 150 days prior to flowering (Best & McIntyre, 1972). It was later

148   found that vernalization increased the rate of flowering in the late flowering accessions

149   (Best & Mc Intyre, 1976). Analyses of F2 progeny between the late and early flowering

150   accessions determined the early flowering phenotype (spring annuals lacking the

151   vernalization requirement) was determined by a single gene, with the late flowering allele

152   being completely dominant (McIntyre & Best, 1978).

153

154   In this report we address the questions concerning the molecular basis for the spring

155   flowering habit in pennycress, whether or not the spring habit has arisen only once or

156   multiple times in nature, and we begin to analyze the geographic distribution of

157   pennycress flowering time variants. These studies were aided by genomic resources that

158   have been developed for pennycress, including a transcriptome and draft genome (Dorn

159   *et al.*, 2013; Dorn *et al.*, 2015) These resources capture the pennycress gene space and

160   greatly aided in the identification of alterations that lead to spring flowering. We show

161   that key mutations in the FLC gene are responsible for spring flowering in many natural

162   pennycress accessions, including one allele of FLC that confers the spring annual growth

163   habit that spread from Europe to North America.

164

165   **Materials and Methods**

166   **1.) *Thlaspi arvense* accessions and F2 population**

167   The spring annual *Thlaspi arvense* line MN108SA was derived from a wild Minnesota

168   population containing both winter and spring annual plants. Five generations of single

169   seed descent was performed on a spring annual plant from this collection and sequenced.

170   A single MN111 plant was also carried through two generations of single seed descent

171   and sequenced.

172   Additional pennycress accessions described here with the 'PI' or 'Ames' prefix were

173   obtained from USDA-GRIN. All pennycress accessions with the 'MN' prefix were

174   collected by Dr. Donald Wyse (University of Minnesota). The spring annual line

175   Spring32 was obtained from Dr. Win Phippen at Western Illinois University under a

176   Materials Transfer Agreement.

177

178   Plants were germinated on moist Berger BM2 germination mix (Berger, Inc.,

179   www.berger.ca), stratified at 4° C for 7 days, and grown in climate-controlled growth

180   chambers at the University of Minnesota (21° C, 16 hour/8 hour day/night cycles at 100

181   micromoles/m^2/s PAR). The MN111 plant sequenced in this analysis was vernalized at

182   six weeks post-germination at 4° C for 30 days in the dark. After vernalization, this plant

183   was returned to the growth chamber conditions described above. The spring annual

184   accessions were not vernalized as they flowered immediately.

185

186   **2.) DNA isolation and Illumina genomic DNA sequencing**

187   DNA was isolated from a single MN108SA and a single MN111 plant using the Omega

188   Mag-Bind Plant DNA kit (Omega Bio-Tek, www.omegabiotek.com) according to the

189   manufacturer's recommended protocol. These DNA samples were submitted to the

190   University of Minnesota Genomics Center for sequencing on the Illumina HiSeq 2000

191   platform (Illumina Inc, www.illumina.com). Sequencing libraries were prepared using

192   the Illumina TruSeq Nano DNA Sample Prep kit with an average library insert size of

193   460 base pairs. The Illumina Universal Adaptor and an Indexed Adaptor (MN111 –

194   Illumina Indexed Adaptor #12 – barcode = CTTGTA, MN108SA – Illumina Indexed

195   Adaptor #11 – barcode = GGCTAC) were used to create the sequencing libraries. Each

196   library was sequenced on a full lane of Illumina HiSeq 2000 (100 base pair, paired-end).

197

198   DNA from the six selected European pennycress accessions was isolated using the

199   Qiagen Plant DNeasy Mini kit (www.qiagen.com) following the manufacturer's

200   recommended protocol. Illumina sequencing libraries were created for each European

201   accession using the Illumina TruSeq Nano kit with the 350 bp insert protocol. These six

202   libraries were sequenced across 1.5 lanes on the Illumina HiSeq 2500 platform (125 base

203   pair, paired-end) with version 4 chemistry.

204

205   All raw sequencing datasets have been deposited in the NCBI Short Read Archive under

206   BioProject PRJNA237017 / SRA Accession SRP036068 (Table S1).

207

208   **3.) Sequencing data quality control, read mapping**

209   Illumina sequencing read quality and contamination were examined using FASTQC

210   (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/). FASTQ files were filtered

211   and trimmed to remove low quality reads and sequencing adaptors using BBDuk (

212   https://sourceforge.net/projects/bbmap/) using the following parameters: ftl=10

213   minlen=50 qtrim=rl trimq=10 ktrim=r k=25 mink=11 hdist=1

214   ref=/bbmap/resources/adapters.fa tpe tbo. Sequencing reads from accessions MN111 and

215   MN108SA had the additional BBDuk flag of ftr=95. The resulting high quality reads

216   from each accession were mapped to the v1 pennycress genome using Bowtie 2 and

217   visualized in CLC Genomics Workbench.

218

219 **4.) Analysis of FLC mutations in MN111 x MN108SA F2 population, global spring**

220 **varieties, and EMS mutant**

221 DNA was isolated from F2 progeny grown in the conditions described above using the

222 Omega Bio-Tek Plant MagBind 96 kit according to the manufacturers recommended

223 protocol. DNA oligos were designed to amplify the 5' end of the pennycress FLC gene

224 for the *flc-A* allele, approximately 100 base pairs upstream of the transcriptional start site

225 (TaFLC_1_Forw: 5' – CCGAGGAAGAAAAAGTAGATAGAGACA – 3',

226 TaFLC_1_Rev: 5' – GAAGCTTAAAGGGGGAAAAAGGAA – 3', Table S2). This

227 amplicon was also able to identify the *flc-D* allele in MN133 and PI633414 and the *flc-α*

228 allele from the EMS-mutagenized population as well. Polymerase Chain Reaction (PCR)

229 was used to amplify this fragment, producing an approximately 450 bp amplicon. New

230 England Biolabs Q5 High-Fidelity PCR Kit with 2x Master Mix was used, with the

231 following thermal cycler conditions: 1.) 98°C for 30 seconds, 2.) 98°C for 10 seconds, 3.)

232 57°C for 20 seconds, 4.) 72°C for 20 seconds, 5.) Go to step #2 34 times , 6.) 72°C for 2

233 minutes, 7.) 4°C hold. Reactions were visualized to confirm amplification of a single

234 band using gel electrophoresis. PCR products were submitted to Beckman Coulter

235 Genomics for PCR product purification and single pass Sanger sequencing. Amplicons

236 were sequenced in both directions using the forward and reverse primers listed above.

237 Sanger sequencing reads were analyzed in CLC Genomics Workbench and aligned

238 against the pennycress MN106 reference genome at the FLC locus to identify sequence

239 variants.

240

241 **5.) Identification of deletion alleles using PCR**

242 A diagnostic PCR test was designed to test for the *flc-B* allele. A primer set

243 (TaFLC_2_Forw – 5' – ATAGTGTGCATCAACTGGTC – 3', and TaFLC_2_Rev – 5'–

244 CGAACCATAGTTCAGAGCTT – 3', Table S2) were designed to amplify a single

245 amplicon overlapping the 456 bp deletion (Fig. S1). In the absence of the *flc-B* allele, a

246 2,088 bp fragment is produced, whereas a 1,632 bp fragment is produced in plants

247 containing the *flc-B* allele (Fig. 2B). New England Biolabs Q5 High-Fidelity PCR Kit

8

248   with 2x Master Mix was used, with the following thermal cycler conditions: 1.) 98°C for

249   30 seconds, 2.) 98°C for 5 seconds, 3.) 63°C for 15 seconds, 4.) 72°C for 60 seconds, 5.)

250   Go to step #2 34 times, 6.) 72°C for 2 minutes, 7.) 4°C hold.

251

252   A diagnostic PCR test was developed to test for the presence or absence of the *flc-C*

253   allele (Fig. 2C, Fig. S2). A primer set (TaFLC_4_Forw: 5'–

254   GCGACGGTGAATATGGAGTTGG – 3', and TaFLC_3_Rev – 5'–

255   GCTAATTTTTCAGCAAATCTCCCG – 3', Table S2) was designed to amplify a 6,598

256   bp amplicon overlapping the 4,724 bp deletion. In plants containing the *flc-C* allele, this

257   region produces a 1,874 bp amplicon, confirming the presence of this deletion. New

258   England Biolabs Q5 High-Fidelity PCR Kit with 2x Master Mix was used, with the

259   following thermal cycler conditions: 1.) 98°C for 30 seconds, 2.) 98°C for 5 seconds, 3.)

260   67°C for 15 seconds, 4.) 72°C for 40 seconds, 5.) Go to step #2 34 times, 6.) 72°C for 2

261   minutes, 7.) 4°C hold. This amplicon was purified using the Qiagen QiaQuick PCR

262   Purification Kit (www.qiagen.com) from accessions MN135, Ames22461, and PI650285

263   and subjected to Sanger sequencing to confirm the exact deletion size and coordinates.

264   Identical 4,724 bp deletions were identified in these three accessions.

265

266   A secondary amplicon was used to confirm the absence of this deletion, with the forward

267   primer (TaFLC_1_Forw: 5' – CCGAGGAAGAAAAAGTAGATAGAGACA – 3', Table

268   S2) lying in the middle of the deleted region (Fig. S2). Successful amplification using this

269   secondary primer set (TaFLC_1_Forw and TaFLC_3_Rev – 5' –

270   GCTAATTTTTCAGCAAATCTCCCG – 3', Table S2) produces a 1,397 bp fragment

271   and confirms the absence of this deleted region. New England Biolabs Q5 High-Fidelity

272   PCR Kit with 2x Master Mix was used, with the following thermal cycler conditions: 1.)

273   98°C for 30 seconds, 2.) 98°C for 5 seconds, 3.) 64°C for 15 seconds, 4.) 72°C for 40

274   seconds, 5.) Go to step #2 34 times, 6.) 72°C for 2 minutes, 7.) 4°C hold.

275

276

277 **Results**

278 *Flowering time phenotypes of winter and spring annual pennycress*

279 While identifying germplasm to be used for the pennycress breeding program, wild

280 isolates of pennycress have been collected from various locations across North America

281 (Table S3). A spring flowering pennycress variant, named MN108SA, was isolated from

282 a mixed population of both spring and winter annual individuals collected from Roseau,

283 Minnesota. Single seed descent for 5 generations showed that the trait was stable and that

284 vernalization was not required for floral induction. MN108SA was crossed to the

285 vernalization-requiring accession MN111. A comparison of MN108SA and MN111

286 seedlings is shown in Fig. 1A. MN108SA shows internode elongation soon after the

287 formation of the first true leaves (Fig. 1A insert), whereas MN111 forms a rosette. As

288 previously reported by McIntyre and Best, 1978, the F1 derived from the cross between

289 MN111 and MN108SA exhibited the dominant winter flowering habit, segregating in the

290 F2 population at 38:12 winter annual:spring annual, suggesting that a single dominant

291 gene is responsible for maintaining the winter annual phenotype.

292

293 Whole genome sequencing of the MN111 and MN108SA parent plants was used to

294 identify key genomic differences underlying the difference in habit between winter

295 annual and spring annual plants. DNA from each parent was sequenced on a full lane of

296 the Illumina HiSeq 2000 platform (100 base pair, paired end). Over 334M sequencing

297 reads were generated per parent (Table S1). A total of 29B high quality (post quality

298 control) were generated per parent, representing >50X coverage of the predicted genome

299 size of 539Mb [1]. Trimmed and filtered reads were mapped to the winter annual MN106
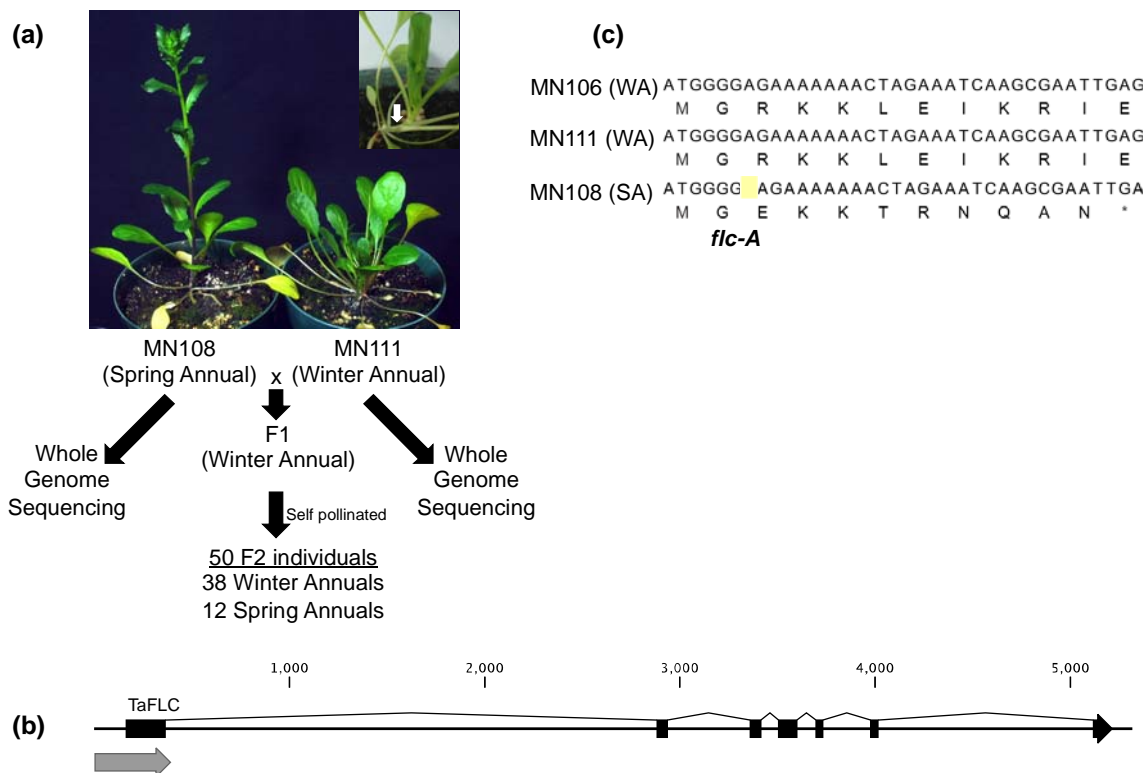
300 draft genome sequence [2].

301

(a) (c)

MN106 (WA) ATGGGGAGAAAAAAACTAGAAATCAAGCGAATTGAG
M G R K K L E I K R I E

MN111 (WA) ATGGGGAGAAAAAAACTAGAAATCAAGCGAATTGAG
M G R K K L E I K R I E

MN108 (SA) ATGGGG AGAAAAAAACTAGAAATCAAGCGAATTGA
M G E K K T R N Q A N *

flc-A

MN108 (Spring Annual) x MN111 (Winter Annual)

Whole Genome Sequencing

F1 (Winter Annual)

Whole Genome Sequencing

Self pollinated

50 F2 individuals
38 Winter Annuals
12 Spring Annuals

(b) TaFLC

302

Fig. 1: Segregation analysis confirms *flc-A* allele confers spring annual phenotype

303

304 (a) Co-grown MN108SA and MN111 pennycress plants. Insert shows immediate

305 internode elongation in the MN108SA spring annual plant. An F2 population was derived

306 from a cross between MN108SA x MN111 and the segregation ratio of spring to winter

307 annual phenotypes was determined. Whole genome sequencing of the parent plants from

308 each background was also conducted. (b) Gene model of pennycress FLOWERING

309 LOCUS C (FLC). Gray arrow below the first exon indicates the position of the amplicon

310 used to genotype the SNP shown in panel C. (c) Nucleotide and predicted peptide

311 sequence of the 5' end of the pennycress FLC gene in winter annual MN106, winter

312 annual MN111, and the spring annual accession MN108SA. The *flc-A* mutation in

313 MN108SA is highlighted in yellow.

314

11

Whole genome sequencing identifies the pennycress *flc-A* allele

Read mappings at the pennycress FRI and FLC loci for MN106, MN111, and MN108SA were examined to identify potential mutations underlying the winter to spring annual transition. Comparison of the pennycress FRI homolog (Ta1.0_26225 on scaffold 1344 of the v1.0 pennycress genome) revealed a single SNP between these three lines, however, this SNP was shared by both the MN111 and MN108SA individual, likely indicating a non-causative effect on the spring annual growth habit (Fig. S3). This SNP was found to cause an amino acid change (Threonine to Serine) at position 553 of the TaFRI predicted peptide (position 2,085 of the gene model - c.2,085A>T).

Within the MN106 draft genome, a single copy of FLC with a high degree of similarity to that of Arabidopsis was identified (Fig. 1B, Fig. S4, Fig. S5). We identified a single guanine base insertion in the 7th position of the FLC coding sequence (*flc-A*) of the spring annual MN108SA accession that was absent in both winter annual accessions. This frameshift mutation caused a predicted TGA stop codon in the 12th codon position (Fig. 1C).

To obtain genetic evidence that this *flc-A allele* was the causal variant responsible for the spring flowering habit in MN108SA, an F2 population from the cross between MN111 and MN108SA was examined. A single F1 individual was self pollinated, and 50 F2 progeny were planted for further analysis. Of these F2 individuals, 38 exhibited a winter annual phenotype and 12 exhibited a spring annual phenotype (Fig. 1A).

PCR primers were developed that amplify a 394 bp amplicon that covers the location of the *flc-A* insertion. Amplicons from all 12 spring annual F2 progeny were sequenced and confirmed homozygous for this *flc-A* allele. Amplicons from a random selection of 14 winter annual F2 progeny were also sequenced. Of the winter annual F2 progeny, 5 were homozygous for the wild type FLC gene while 9 were heterozygous. A chi square

343   analysis ($\chi 2$=0.026667, p-value = 0.8703) indicates that the data fits a 1:2:1 segregation

344   pattern that is predicted for a single recessive mutation in this F2 population.

345

346   **Identifying a novel EMS-induced allele of FLC in pennycress**

347   As a component of ongoing work to domesticate pennycress, an EMS-mutagenized

348   MN106 population has been generated (Sedbrook, 2014). This winter flowering

349   population has been screened to identify early flowering variants. In the course of this

350   screen, a new spring flowering time mutant was identified (Fig. S6). This line did not

351   require vernalization and flowered immediately upon germination. Whole genome

352   sequencing of this mutant revealed a novel mutation in the first exon of FLC (*c.52C>T*),

353   resulting in an immediate stop codon and truncated predicted protein (Fig. S6), which

354   was designated *flc-α*. The above segregation data along with the finding of two new

355   independent spring flowering mutants with FLC mutations confirms the causative role of

356   FLC mutations in the spring flowering habit in pennycress.
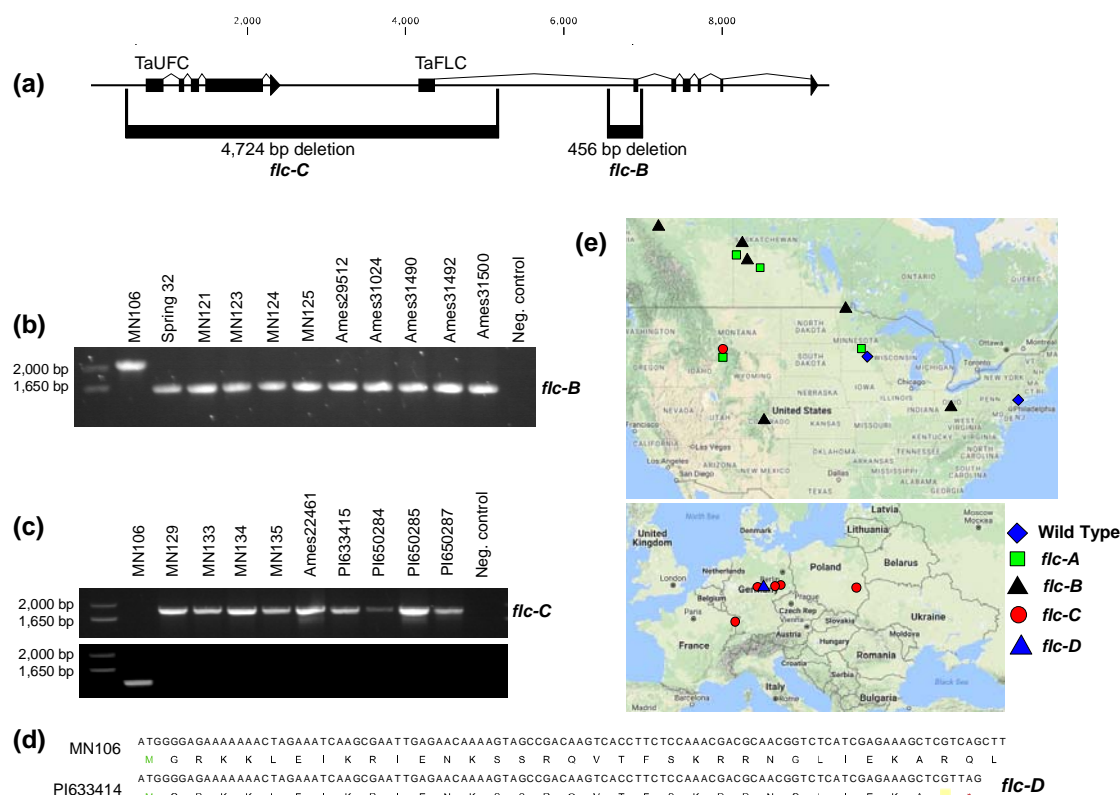
357

358

359

360

361

Fig. 2: Identification of *flc-B* and *flc-C* alleles of FLC

(a) Expanded gene model of pennycress FLC region, including the upstream TaUFC gene model. Locations of the *flc-B* and *flc-C* alleles are shown below the gene model. (b) Agarose gel of the diagnostic polymerase chain reaction (PCR) results testing for the *flc-B* allele across the shown pennycress accessions. The wild type MN106 allele generates a 2,088 bp amplicon, where as the *flc-B* allele produces a 1,632 bp amplicon (Fig S1). (c) Agarose gel of the diagnostic polymerase chain reaction (PCR) results testing for the *flc-C* allele across the shown pennycress accessions. Amplification in the top row confirms the presence of the deletion, whereas amplification in the second row confirms absence of the deletion. The sequences of primers used are shown in Table S2, and primer locations and amplicon strategy is shown in Fig. S2. Results for accession PI650286 (Groitzsch, Saxony, Germany) are not shown, but was confirmed to have the *flc-C* allele. (d)

14

374   Nucleotide and predicted peptide sequence of FLC in MN106 and PI633414 (Wachstedt,

375   Germany), with the *flc-D* allele highlighted in yellow. (e) Global distribution of the four

376   spring annual alleles of FLC in pennycress accessions tested in this study (listed in Table

377   S3).

378

379   **Global distribution of three additional spring annual alleles of FLC**

380   With the genetic confirmation that the *flc-A* allele was indeed the causal variant

381   responsible for the spring annual growth habit in MN108SA, we next examined the

382   distribution of this allele in spring annual accessions from around the world. We screened

383   a total of 35 spring annual accessions, which were obtained either from the University of

384   Minnesota collections or GRIN. Of these, 23 were confirmed as pure spring annual

385   accessions in both field and growth chamber experiments, with no winter annual

386   segregation nor contaminating winter annual seed (Table S3). Surprisingly, we only

387   identified the *flc-A* allele in three other accessions, all from North America (MN131 from

388   Montana, USA and Ames31489 and Ames31491 from Canada).

389

390   As there were a limited number of representative spring annual accessions in our

391   collection containing the *flc-A* allele, we utilized a PCR-based approach to determine

392   whether there were additional mutations in FLC underlying the spring annual growth

393   habit. As we had already amplified and sequenced the first exon of FLC in all of our

394   spring annual accessions, we moved next to the second exon. Interestingly, we were

395   unable to amplify the second exon in several spring annual accessions. Thus, we

396   expanded our PCR experiments into the flanking introns around exon 2, which lead to the

397   discovery of a second allele of FLC in spring annual lines. This mutation was

398   characterized by a 456 bp deletion that encompasses the entire second exon of FLC,

399   which was named *flc-B*. This allele was identified in 9 spring annual accessions

400   distributed across the United States and Canada, along with the Spring32 accession from

401   the Western Illinois University breeding program (Fig. 2B and 2E, Table S3) (Sedbrook,

402   2014).

403

404   Having not yet identified a candidate spring annual FLC allele in any accession from

405   Europe, we again employed whole genome sequencing of the 5 publicly available spring

406   annual pennycress accessions (collected in Germany, Poland, and France). Using this

407   approach, we were able to quickly identify a third allele, characterized by a 4,724 bp

408   deletion (*flc-C*), which removes the first intron of FLC along with the upstream gene

409   (gene model Ta000918), a homolog of the Arabidopsis gene UPSTREAM OF FLC (UFC,

410   AT5G10150). In Arabidopsis, UFC and FLC are co-regulated, however, UFC is not

411   believed to influence the vernalization response (Finnegan *et al.*, 2004; Andersson *et al.*,

412   2008). A diagnostic amplicon experiment was designed to screen the remaining

413   unknown spring annual accessions for the *flc-C* allele. This allele was identified in four

414   accessions collected in Howard Springs, Montana, USA, along with three accessions

415   from Germany, as well as accessions from Poland and France (Fig. 2C, 2E, Supplemental

416   Table 3).

417

418   A fourth allele, *flc-D*, was identified through whole genome sequencing of the spring

419   annual accession PI633414 (Germany). This allele is characterized by a single base pair

420   change at the 100th position of the CDS and caused a precocious TAG stop codon in the

421   predicted protein sequence (Fig. 2D). As the position of this mutation was covered by the

422   394 bp amplicon used to identify the *flc-A* allele, we re-examined the Sanger reads from

423   all of the accessions sequenced in our search for accessions containing the *flc-A* allele.

424   We did not identify any additional accessions containing the *flc-D* allele.

425

426

427

428

429

430

Discussion

Use of NGS to quickly identify mutations in FLC

At the onset of this experiment, the predicted costs associated with PCR-based cloning and sequencing of all candidate genes was assessed. This analysis indicated that it is now less expensive to use the whole genome sequencing (WGS) approach to first identify a candidate locus in the parent accessions, and then proceed to test F2 progeny via PCR and Sanger sequencing. Here we report the discovery of four natural alleles of FLC that confer the spring annual phenotype in pennycress. The added value with the WGS approach here is that we now also have genome-wide markers for the two parents of the F2 mapping population (MN111 and MN108SA), which is being used to develop a linkage map via Genotype by Sequencing (Elshire *et al.,* 2011; Poland & Rife, 2012). Additionally, as the network controlling flowering time is known to be extremely complex, the gene variants in each of these parents that may have a minor effect on vernalization and flowering provide a wealth of untapped information for later investigation.

Global distribution of FLC alleles and the introduction to North America

Pennycress is native to Europe and Asia. Thought to have been introduced to North America as early as the establishment of Fort Detroit (Michigan, USA) in 1701, pennycress became widely distributed throughout Canada and the United States by the early 1900s (Best & McIntyre, 1975). Despite the limited number (35) of pure spring annual accessions tested in this study, we can begin examining the origins of each of the spring annual alleles of FLC identified here.

As the species originated in Eurasia and only the *flc-B and flc-D* alleles were identified in the seven spring annual European accessions tested here, we hypothesize that these alleles originated prior to introduction to North America (Fig. 2E). As the accessions collected near Howard Spring, Montana, USA (MN129, MN133, MN134, MN135) contained one of these 'European' alleles of FLC (*flc-C*), this suggests that this allele

17

460   may have been introduced into North America a limited number of times, potentially

461   from plants near Saxony or Thuringia, Germany (Fig. 2E). However, it is possible that

462   the *flc-D* allele is present in populations in North America, but was not identified in the

463   accessions represented in this experiment.

464

465   Similarly, as we did not identify European accessions containing the *flc-A* or *flc-B* alleles,

466   we predict that these alleles may have originated after the introduction of pennycress to

467   North America and this allele is unique to North American accessions. It is possible that

468   there are in fact European pennycress populations that contain the *flc-A* allele, but with

469   the limited European germplasm availability, were not found in this study. In this

470   scenario, it is feasible that a mixed seed lot containing *flc-A* and *flc-B* alleles were

471   introduced to Montana in a single event, and plants with the *flc-A* and *flc-B* alleles spread

472   significantly more compared to those with the *flc-D* allele.

473

474   The accessions collected near Howard Spring, Montana (MN129, MN131, MN133,

475   MN134, MN135) are an interesting case study of these hypotheses, especially

476   considering the findings of the *flc-A* allele in MN131, the *flc-B* allele in MN129, MN133,

477   MN134, and MN135, and a lack of identification of European accessions with the *flc-A*

478   allele. The above scenarios could be investigated further in these accessions collected

479   from Montana using genome-wide markers from either genotype-by-sequencing or WGS.

480   As we have already generated WGS data for five of the European accessions in this study,

481   sequencing of these accessions from Montana and a comparison of shared

482   polymorphisms may help in determining the relatedness between the North American and

483   European lines, and may point to the exact origins of these alleles of FLC in pennycress.

484   Again, with the limited number of pure spring annual accessions available, future studies

485   should also focus on the collection of a more comprehensive set of globally distributed

486   lines, including both spring and winter annuals.

487

488

489   **Towards an understanding of flowering time control in pennycress**

490   Our previous efforts to develop genomic resources for the domestication of pennycress

491   identified likely homologs controlling flowering time and the vernalization response via

492   DNA and RNA sequence homology (Dorn *et al.*, 2013; Dorn *et al.*, 2015). Here we

493   present the first sequence-supported genetic information on the underlying mechanisms

494   controlling flowering time in pennycress. While the four natural alleles of FLC reported

495   here are sufficient to confer the spring annual phenotype, there is likely a host of

496   interacting genes and epigenetic factors also essential for the rapid flowering seen in the

497   spring annual accessions investigated here. Of particular interest is FRIGIDA and

498   members of the 'FRIGIDA Complex' (FRI-C), including FRIGIDA LIKE 1 (FRL1),

499   FRIGIDA-ESSENTIAL 1 (FES1), SUPPRESSOR OF FRIGIDA4 (SUF4), and FLC

500   EXPRESSOR (FLX) (Choi *et al.*, 2011). The FRIGIDA Complex acts as a transcriptional

501   activation complex on FLC expression through a diverse range of functions of each

502   complex member (Choi *et al.*, 2011). While no loss of function FRIGIDA alleles were

503   identified in the accessions examined here, it remains possible that an FLC-independent

504   path to the spring annual phenotype exists in light of the limited availability of globally

505   distributed germplasm. The interaction between the vernalization response, photoperiodic,

506   and autonomous pathways are also of great interest, as allelic variation and unique

507   combinations of alleles from each pathway can contribute to quantitative variation in

508   flowering time and lead towards the goal of developing an elite winter annual pennycress

509   variety that flowers rapidly in the spring to fit within the short growing season of

510   northern climates.

511

512

513

514

515

516

517

518   **Acknowledgements**

519   The authors would like to thank Ryan Emenecker, Nicole Folstad, Matthew Ott, and

520   Ratan Chopra for their assistance with experiments and for feedback on the manuscript.

521   This material is based upon work supported by the National Science Foundation Graduate

522   Research Fellowship under Grant No. 00006595 to KMD. Any opinions, findings, and

523   conclusions or recommendations expressed in this material are those of the authors and

524   do not necessarily reflect the views of the National Science Foundation. This project is

525   funded by the USDA National Institute of Food Agriculture - Institute of Bioenergy,

526   Climate and Environment, competitive grant no. 2014-67009-22305.

527

528

529   **Author Contributions**

530   KMD and MDM conceived and planned the study. EBJ, ED, and MDM constructed the

531   F2 mapping population. EBJ and ED identified the FLC-α allele. KMD designed and

532   performed PCR experiments, whole genome sequencing and analyses, and comparative

533   analyses. DLW collected all 'MN' designated pennycress accessions used in this study.

534   KMD and MDM wrote the manuscript with feedback from all the co-authors.

535

536

537

538

539

540

541

542

543

544

545

546

References

Amasino RM. 2005. Vernalization and flowering time. *Curr Opin Biotechnol* **16**(2): 154-158.

Andersson CR, Helliwell CA, Bagnall DJ, Hughes TP, Finnegan EJ, Peacock WJ, Dennis ES. 2008. The FLX gene of Arabidopsis is required for FRI-dependent activation of FLC expression. *Plant and cell physiology* **49**(2): 191-200.

Beilstein MA, Al-Shehbaz IA, Mathews S, Kellogg EA. 2008. Brassicaceae phylogeny inferred from phytochrome A and ndhF sequence data: tribes and trichomes revisited. *Am J Bot* **95**(10): 1307-1327.

Best K, Mc Intyre G. 1976. Studies on the flowering of Thlaspi arvense L. III. The influence of vernalization under natural and controlled conditions. *Botanical Gazette*: 121-127.

Best K, McIntyre G. 1972. Studies on the Flowering Time of Thlaspi arvense L. I. The Influence of some Environmental and Genetic Factors. *Botanical Gazette* **133**(4): 454 - 459.

Best K, McIntyre G. 1975. The Biology of Canadian Weeds 9. Thlaspi arvense L. . *Canadian Journal of Plant Science* **55**: 279-292.

Burn J, Smyth D, Peacock W, Dennis E. 1993. Genes conferring late flowering inArabidopsis thaliana. *Genetica* **90**(2-3): 147-155.

Choi K, Kim J, Hwang HJ, Kim S, Park C, Kim SY, Lee I. 2011. The FRIGIDA complex activates transcription of FLC, a strong flowering repressor in Arabidopsis, by recruiting chromatin modification factors. *Plant Cell* **23**(1): 289-303.

Clarke JH, Dean C. 1994. Mapping FRI, a locus controlling flowering time and vernalization response in Arabidopsis thaliana. *Mol Gen Genet* **242**(1): 81-89.

Coustham V, Li P, Strange A, Lister C, Song J, Dean C. 2012. Quantitative modulation of polycomb silencing underlies natural variation in vernalization. *Science* **337**(6094): 584-587.

Crowley JG. 1999. *Evaluation of Camelina sativa as an alternative oilseed crop*: Teagasc.

Csorba T, Questa JI, Sun Q, Dean C. 2014. Antisense COOLAIR mediates the coordinated switching of chromatin states at FLC during vernalization. *Proceedings of the National Academy of Sciences* **111**(45): 16160-16165.

Dorn KM, Fankhauser JD, Wyse DL, Marks MD. 2013. De novo assembly of the pennycress (Thlaspi arvense) transcriptome provides tools for the development of a winter cover crop and biodiesel feedstock. *Plant J* **75**(6): 1028-1038.

Dorn KM, Fankhauser JD, Wyse DL, Marks MD. 2015. A draft genome of field pennycress (Thlaspi arvense) provides tools for the domestication of a new winter biofuel crop. *DNA Res* **22**(2): 121-131.

Elshire RJ, Glaubitz JC, Sun Q, Poland JA, Kawamoto K, Buckler ES, Mitchell SE. 2011. A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. *PLoS One* **6**(5): e19379.

589 **Evangelista RL, Isbell TA, Cermak SC. 2012.** Extraction of pennycress (Thlaspi
590   arvense L.) seed oil by full pressing. *Industrial Crops and Products* **37**(1): 76-81.
591 **Fan J, Shonnard DR, Kalnes TN, Johnsen PB, Rao S. 2013.** A life cycle assessment of
592   pennycress (Thlaspi arvense L.) -derived jet fuel and diesel. *Biomass and*
593   *Bioenergy* **55**: 87-100.
594 **Finnegan EJ, Dennis ES. 2007.** Vernalization-induced trimethylation of histone H3
595   lysine 27 at FLC is not maintained in mitotically quiescent cells. *Current Biology*
596   **17**(22): 1978-1983.
597 **Finnegan EJ, Sheldon CC, Jardinaud F, Peacock WJ, Dennis ES. 2004.** A cluster of
598   Arabidopsis genes with a coordinate response to an environmental stimulus.
599   *Current Biology* **14**(10): 911-916.
600 **Franzke A, Lysak MA, Al-Shehbaz IA, Koch MA, Mummenhoff K. 2011.** Cabbage
601   family affairs: the evolutionary history of Brassicaceae. *Trends Plant Sci* **16**(2):
602   108-116.
603 **Greb T, Mylne JS, Crevillen P, Geraldo N, An H, Gendall AR, Dean C. 2007.** The
604   PHD finger protein VRN5 functions in the epigenetic silencing of Arabidopsis
605   FLC. *Current Biology* **17**(1): 73-78.
606 **Hojilla-Evangelista MP, Evangelista RL, Isbell TA, Selling GW. 2013.** Effects of
607   cold-pressing and seed cooking on functional properties of protein in pennycress
608   (Thlaspi arvense L.) seed and press cakes. *Industrial Crops and Products* **45**: 223-
609   229.
610 **Jiang D, Gu X, He Y. 2009.** Establishment of the winter-annual growth habit via
611   FRIGIDA-mediated histone methylation at FLOWERING LOCUS C in
612   Arabidopsis. *Plant Cell* **21**(6): 1733-1746.
613 **Johanson U, West J, Lister C, Michaels S, Amasino R, Dean C. 2000.** Molecular
614   analysis of FRIGIDA, a major determinant of natural variation in Arabidopsis
615   flowering time. *Science* **290**(5490): 344-347.
616 **Jung C, Muller AE. 2009.** Flowering time control and applications in plant breeding.
617   *Trends Plant Sci* **14**(10): 563-573.
618 **Kim DH, Doyle MR, Sung S, Amasino RM. 2009.** Vernalization: winter and the timing
619   of flowering in plants. *Annu Rev Cell Dev Biol* **25**: 277-299.
620 **Marquardt S, Raitskin O, Wu Z, Liu F, Sun Q, Dean C. 2014.** Functional
621   consequences of splicing of the antisense transcript COOLAIR on FLC
622   transcription. *Mol Cell* **54**(1): 156-165.
623 **McIntyre G, Best K. 1978** Studies on the Flowering of Thlaspi arvense IV. Genetic and
624   Ecological Differences between Early and Late Flowering Strains *Botanical*
625   *Gazette* **139**(2): 190-195.
626 **Michaels SD, Amasino RM. 1999.** FLOWERING LOCUS C encodes a novel MADS
627   domain protein that acts as a repressor of flowering. *Plant Cell* **11**(5): 949-956.
628 **Michaels SD, Amasino RM. 2001.** Loss of FLOWERING LOCUS C activity eliminates
629   the late-flowering phenotype of FRIGIDA and autonomous pathway mutations
630   but not responsiveness to vernalization. *Plant Cell* **13**(4): 935-941.

Michaels SD, Bezerra IC, Amasino RM. 2004. FRIGIDA-related genes are required for the winter-annual habit in Arabidopsis. *Proc Natl Acad Sci U S A* **101**(9): 3281-3285.

Michaels SD, He Y, Scortecci KC, Amasino RM. 2003. Attenuation of FLOWERING LOCUS C activity as a mechanism for the evolution of summer-annual flowering behavior in Arabidopsis. *Proc Natl Acad Sci U S A* **100**(17): 10102-10107.

Moser BR, Knothe G, Vaughn SF, Isbell TA. 2009. Production and Evaluation of Biodiesel from Field Pennycress (Thlaspi arvense L.) Oil. *Energy & Fuels* **23**(8): 4149-4155.

Phippen WB, Phippen ME. 2012. Soybean Seed Yield and Quality as a Response to Field Pennycress Residue. *Crop Science* **52**(6): 2767.

Poland JA, Rife TW. 2012. Genotyping-by-Sequencing for Plant Breeding and Genetics. *Plant Genome* **5**(3): 92-102.

Searle I, He Y, Turck F, Vincent C, Fornara F, Kröber S, Amasino RA, Coupland G. 2006. The transcription factor FLC confers a flowering response to vernalization by repressing meristem competence and systemic signaling in Arabidopsis. *Genes & Development* **20**(7): 898-912.

Sedbrook JP, WB; Marks, MD 2014. New approaches to facilitate rapid domestication of a wild plant to an oilseed crop: example pennycress (Thlaspi arvense L.). *Plant Sci* **227**: 122-132.

Selling G, Evangelista R, Isbell T, Evangelista MH. 2011. Extraction of proteins from pennycress. *Abstracts of Papers of the American Chemical Society* **241**.

Shindo C, Aranzana MJ, Lister C, Baxter C, Nicholls C, Nordborg M, Dean C. 2005. Role of FRIGIDA and FLOWERING LOCUS C in determining variation in flowering time of Arabidopsis. *Plant Physiol* **138**(2): 1163-1173.

Simpson GG, Dean C. 2002. Arabidopsis, the Rosetta stone of flowering time? *Science* **296**(5566): 285-289.

Stinchcombe JR, Weinig C, Ungerer M, Olsen KM, Mays C, Halldorsdottir SS, Purugganan MD, Schmitt J. 2004. A latitudinal cline in flowering time in Arabidopsis thaliana modulated by the flowering time gene FRIGIDA. *Proc Natl Acad Sci U S A* **101**(13): 4712-4717.

Sung S, He Y, Eshoo TW, Tamada Y, Johnson L, Nakahigashi K, Goto K, Jacobsen SE, Amasino RM. 2006. Epigenetic maintenance of the vernalized state in Arabidopsis thaliana requires LIKE HETEROCHROMATIN PROTEIN 1. *Nat Genet* **38**(6): 706-710.

Swiezewski S, Liu F, Magusin A, Dean C. 2009. Cold-induced silencing by long antisense transcripts of an Arabidopsis Polycomb target. *Nature* **462**(7274): 799.

Tadege M, Sheldon CC, Helliwell CA, Stoutjesdijk P, Dennis ES, Peacock WJ. 2001. Control of flowering time by FLC orthologues in Brassica napus. *The Plant Journal* **28**(5): 545-553.

Warwick SI, Francis, A., Susko, D.J. . 2002 The biology of Canadian weeds. 9. Thlaspi arvense L. (updated) *Canadian Journal of Plant Science* **82**: 803-823.

673 **Wu J, Wei K, Cheng F, Li S, Wang Q, Zhao J, Bonnema G, Wang X. 2012.** A
674   naturally occurring InDel variation in BraA. FLC. b (BrFLC2) associated with
675   flowering time variation in Brassica rapa. *BMC Plant Biol* **12**(1): 151.
676 **Yang H, Howard M, Dean C. 2014.** Antagonistic roles for H3K36me3 and H3K27me3
677   in the cold-induced epigenetic switch at Arabidopsis FLC. *Current Biology*
678   **24**(15): 1793-1797.

679

680

681 **Supporting Information Brief Legends**

682 **Fig. S1 -** PCR strategy for identifying *flc-B* allele

683 **Fig. S2 -** PCR strategy for identifying *flc-C* allele

684 **Fig. S3 -** Peptide alignment of Brassicaceae FRIGIDA predicted peptides and pennycress

685 FRI gene model

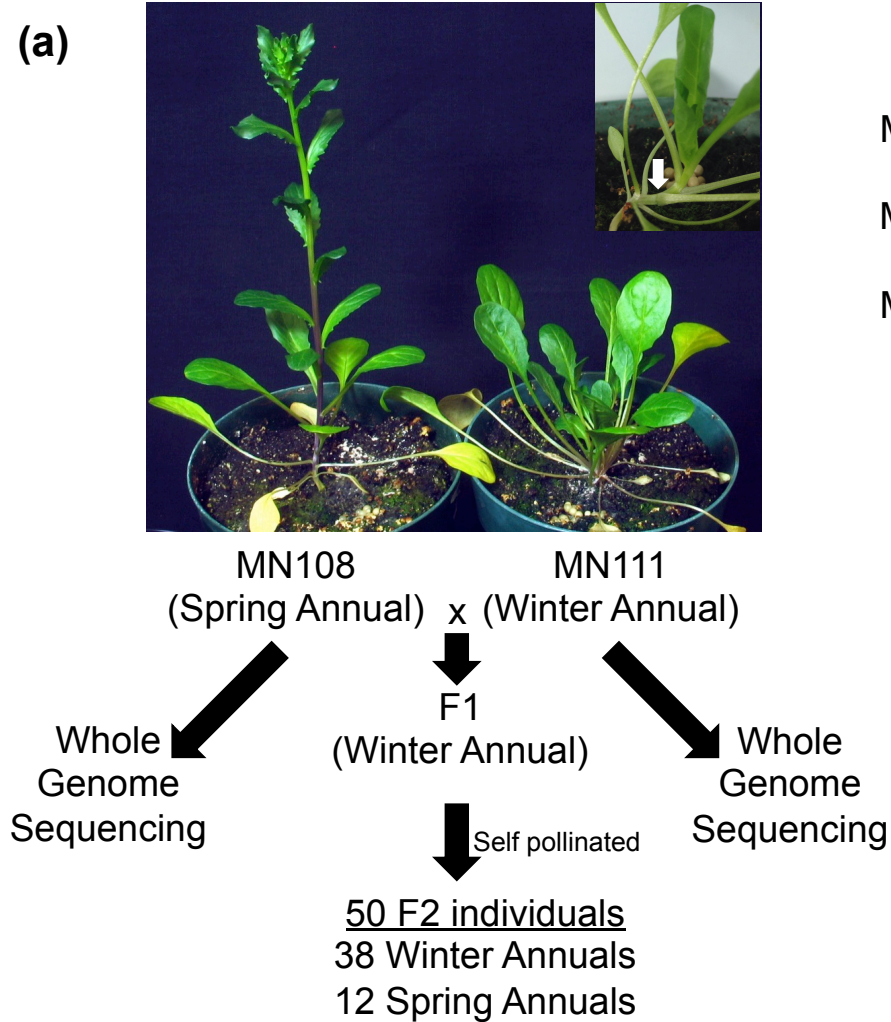686 **Fig. S4 -** Peptide alignment of Brassicaceae FLOWERING LOCUS C predicted peptides

687 **Fig. S5 -** Genomic DNA alignment of pennycress FLOWER LOCUS C wild type and *flc-*

688 *a* allele

689 **Fig. S6 -** Identification and characterization of pennycress flc-α

690 **Table S1 -** Summary of whole genome sequencing reads of pennycress accessions used

691 in this study

692 **Table S2 -** Primer sequences used in PCR experiments used in this study

693 **Table S3 -** Summary of pennycress accessions and the corresponding alleles of FLC as

694 identified in this study

695

(a) MN108 (Spring Annual) x MN111 (Winter Annual)

MN108 (Spring Annual) → Whole Genome Sequencing

F1 (Winter Annual)

MN111 (Winter Annual) → Whole Genome Sequencing

Self pollinated

50 F2 individuals
38 Winter Annuals
12 Spring Annuals
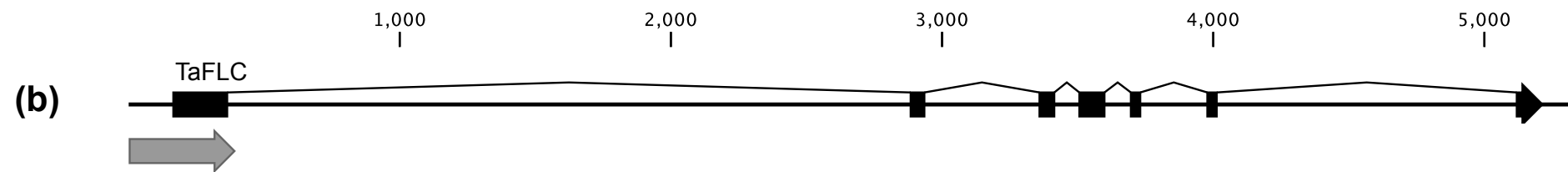
(c)

```
MN106 (WA)   ATGGGGAGAAAAAAACTAGAAATCAAGCGAATTGAG
              M   G   R   K   K   L   E   I   K   R   I   E

MN111 (WA)   ATGGGGAGAAAAAAACTAGAAATCAAGCGAATTGAG
              M   G   R   K   K   L   E   I   K   R   I   E

MN108 (SA)   ATGGGGGAGAAAAAAACTAGAAATCAAGCGAATTGA
              M   G   E   K   K   T   R   N   Q   A   N   *
                            flc-A
```
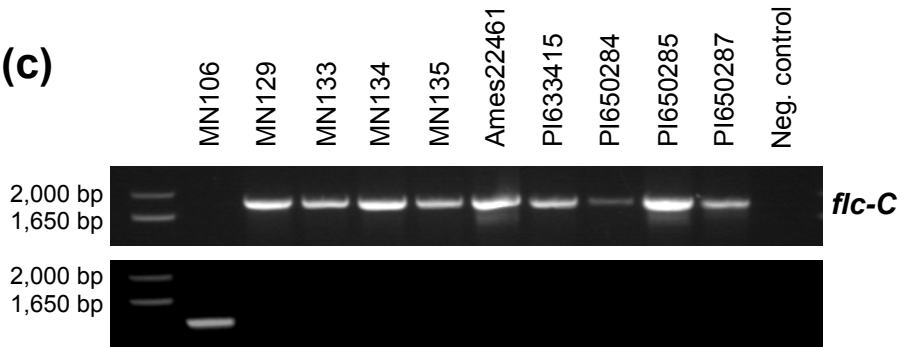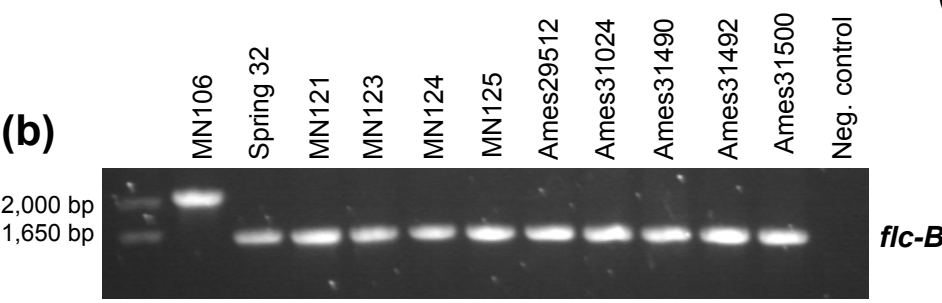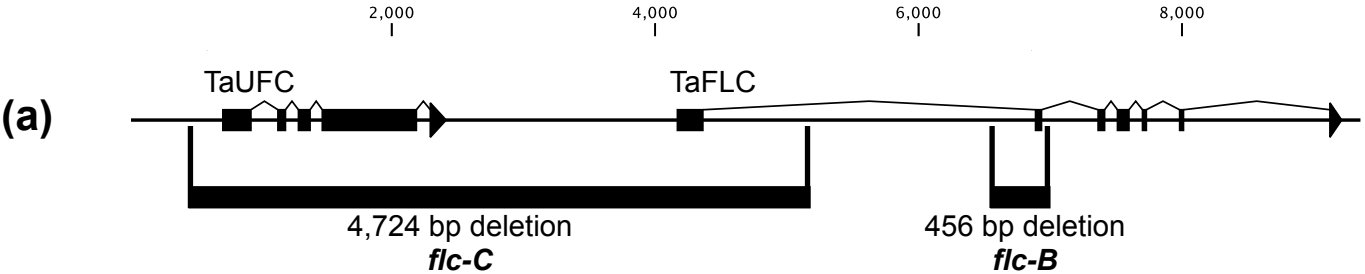
(b)

(a)

TaUFC   TaFLC

2,000   4,000   6,000   8,000

4,724 bp deletion
flc-C

456 bp deletion
flc-B

(b)

MN106 · Spring 32 · MN121 · MN123 · MN124 · MN125 · Ames29512 · Ames31024 · Ames31490 · Ames31492 · Ames31500 · Neg. control

2,000 bp
1,650 bp

flc-B

(c)

MN106 · MN129 · MN133 · MN134 · MN135 · Ames22461 · PI633415 · PI650284 · PI650285 · PI650287 · Neg. control

2,000 bp
1,650 bp

flc-C

2,000 bp
1,650 bp

(e)

Wild Type
flc-A
flc-B
flc-C
flc-D

(d)

MN106
ATGGGGAGAAAAAACTAGAAATCAAGCGAATTGAGAACAAAAGTAGCCGACAAGTCACCTTCTCCAAACGACGCAACGGTCTCATCGAGAAAGCTCGTCAGCTT
M   G R K K L E I K R I E N K S S R Q V T F S K R R N G L I E K A R Q L

PI633414
ATGGGGAGAAAAAACTAGAAATCAAGCGAATTGAGAACAAAAGTAGCCGACAAGTCACCTTCTCCAAACGACGCAACGGTCTCATCGAGAAAGCTCGTTAG
M   G R K K L E I K R I E N K S S R Q V T F S K R R N G L I E K A   R   *

flc-D