ORIGINAL RESEARCH

American Society of Plant Biologists | S·E·B SOCIETY FOR EXPERIMENTAL BIOLOGY | WILEY

Spring flowering habit in field pennycress (*Thlaspi arvense*) has arisen multiple independent times

Kevin M. Dorn[1] | Evan B. Johnson[1] | Erin C. Daniels[1] | Donald L. Wyse[2] | Michael D. Marks[1]

[1]Department of Plant and Microbial Biology, University of Minnesota, Saint Paul, Minnesota

[2]Department of Agronomy and Plant Genetics, University of Minnesota, Saint Paul, Minnesota

Correspondence
Michael D. Marks, University of Minnesota, Department of Plant Biology, St Paul, MN.
Email: marks004@umn.edu

Abstract

Field pennycress (*Thlaspi arvense* L.) is currently being developed as a new cold-tolerant oilseed crop. In natural populations, pennycress, like many Brassicaceae relatives, can exhibit either a winter or spring annual phenotype. Pennycress is a diploid relative of *Arabidopsis thaliana*, a model species that has been used to study many adaptive phenotypes, including flowering time and developmental timing. In Arabidopsis and other Brassicaceae species, mutations in negative regulators of flowering, including FLOWERING LOCUS C and FRIGIDA can cause the transition to a spring annual habit. The genetics underlying the difference between spring and winter annual pennycress lines are currently unknown. Here, we report the identification of four natural alleles of FLC in pennycress that confer a spring annual growth habit identified through whole genome sequencing, cosegregation analyses, and comparative genomics. The global distribution of these spring annual alleles of FLC suggests that the spring annual growth habit has arisen on several independent occasions. The two spring annual FLC alleles present in European accessions were only identified in North American accessions collected in southern Montana, which indicates accessions harboring these two alleles were introduced to North America, likely after pennycress became a widespread species on the continent. These findings provide new information on the natural history of the introduction and spread of spring annual pennycress accessions from Europe into North America. At the molecular level, these findings are important for the ongoing development of pennycress as a winter annual crop. An enhanced understanding of the regulation of flowering in this species should allow for the fine-tuning of flowering in commercial varieties.

KEYWORDS
flowering locus C, pennycress, spring annual, *Thlaspi arvense*, whole genome sequencing, winter annual

1 | INTRODUCTION

Thlaspi arvense L. (field pennycress, pennycress herein) is currently a target for domestication as a new, cold-hardy winter oilseed crop that can fit within the corn/soybean rotation in the Midwestern United States (Dorn, Fankhauser, Wyse, & Marks, 2013, 2015; Sedbrook, Phippen, & Marks, 2014). As a winter annual crop, pennycress can be planted in the late summer to early autumn, either into standing corn or immediately after corn harvest. The pennycress crop establishes a robust vegetative cover prior to winter, providing important ecosystem services such as limiting soil erosion and nutrient run-off. In the spring, pennycress will flower and set seed, yielding upwards of 1,600 pounds per acre of oil-rich seed in time for planting a crop of soybean (Phippen & Phippen, 2012; Sedbrook et al., 2014). Pennycress seed is high in oils that can be readily converted to biodiesel or jet fuel, with the remaining high-protein seed meal having potential as animal feed or feedstock for industrial uses (Evangelista, Isbell, & Cermak, 2012; Fan, Shonnard, Kalnes, Johnsen, & Rao, 2013; Hojilla-Evangelista, Evangelista, Isbell, & Selling, 2013; Moser, Knothe, Vaughn, & Isbell, 2009; Selling, Evangelista, Isbell, & Evangelista, 2011).

Pennycress is a member of the Thlaspideae, a tribe of the Brassicaceae (Beilstein, Al-Shehbaz, Mathews, & Kellogg, 2008). The Brassicaceae family is divided into three lineages. Pennycress resides in lineage II, along with the Brassica genus, including the economically important oilseed crops *Brassica rapa* and *Brassica napus*, but not the model species *Arabidopsis thaliana* nor *Capsella rubella*, which are members of lineage I (Franzke, Lysak, Al-Shehbaz, Koch, & Mummenhoff, 2011). Pennycress is native to Eurasia and is considered naturalized to most temperate to subarctic regions throughout the northern hemisphere, including all of the United States except for Hawaii and Alabama, all provinces of Canada, as well as being considered naturalized in the southern hemisphere, including Australia, New Zealand, and Argentina (Warwick, Francis, & Susko, 2002). Pennycress exhibits numerous traits that make it an attractive winter rotation crop (Sedbrook et al., 2014). Of particular importance in the upper Midwest is the existence of pennycress lines that complete their life cycle rapidly enough to fit within the corn/soybean rotation. Central to this rapid spring development of pennycress are its underlying growth habits. In wild populations, there are both winter and spring annual pennycress (Best & McIntyre, 1972; McIntyre & Best, 1978), similar to many other Brassicaceae species such as *Arabidopsis thaliana* (Shindo et al., 2005; Stinchcombe et al., 2004), *Brassica rapa* (Wu et al., 2012), *Camelina sativa* (Crowley, 1999), and *Brassica napus* (Tadege et al., 2001).

Throughout the decades of research on basic developmental questions in Arabidopsis, and the expanding translational research in other Brassica crops, the underlying molecular mechanisms controlling flowering time have been identified in many of these species (Amasino, 2005; Jung & Muller, 2009; Kim, Doyle, Sung, & Amasino, 2009; Simpson & Dean, 2002). In wild accessions of Arabidopsis, only a handful of mutations are responsible for variation in flowering time and vernalization requirement (Burn, Smyth, Peacock, & Dennis, 1993; Clarke & Dean, 1994; Johanson et al., 2000). Most notably, allelic variation in two key negative regulators, FLOWERING LOCUS C (FLC) and/or FRIGIDA (FRI), underlies the key difference between spring and winter annual plants (Jiang, Gu, & He, 2009; Johanson et al., 2000; Michaels & Amasino, 2001; Michaels, Bezerra, & Amasino, 2004; Michaels, He, Scortecci, & Amasino, 2003; Shindo et al., 2005). FLC encodes a MADS box transcription factor (Michaels & Amasino, 1999), and inhibits flowering prior to vernalization by repressing the expression of FLOWERING LOCUS T (FT) (Searle et al., 2006). In Arabidopsis, *flc* null mutations eliminate the vernalization requirement and impart the spring annual, rapid-flowering phenotype (Michaels & Amasino, 1999). Allelic variation within FRI also impacts flowering time and the vernalization requirement. Similar to FLC, Arabidopsis accessions harboring loss of function mutations in FRI flower rapidly without vernalization (Johanson et al., 2000; Shindo et al., 2005).

The expression of FLC is positively regulated by FRI (Michaels & Amasino, 1999), thus, mutations in either of these vernalization-responsive negative regulators can lead to a loss of vernalization requirement and a spring annual growth habit (Michaels & Amasino, 1999, 2001). The vernalization signal provided by the cold of winter removes the repression on the transition to flowering through the epigenetic silencing of FLC. Specifically, vernalization increases histone 3K27 trimethylation (H3K27me3) at FLC chromatin, reducing transcriptional activity (Coustham et al., 2012; Finnegan & Dennis, 2007; Greb et al., 2007; Sung et al., 2006; Yang, Howard, & Dean, 2014). The vernalization-induced silencing of FLC releases the repression on the transition to flowering, which permits the transition to reproductive growth (Searle et al., 2006). Additionally, the roles of the COOLAIR long noncoding antisense transcripts transcribed from the AtFLC locus have been revealed as a key regulatory component in the cold-induced regulation of FLC (Csorba, Questa, Sun, & Dean, 2014; Marquardt et al., 2014; Swiezewski, Liu, Magusin, & Dean, 2009).

Although this extensive understanding of the molecular genetic pathways controlling flowering time and vernalization in Arabidopsis has informed similar studies in Brassica relatives, little is known about the underlying mechanisms controlling flowering time variation in pennycress. Different accessions of pennycress have been reported to either act as early flowering or late flowering, with late flowering accessions growing as rosettes for as long as 150 days prior to flowering (Best & McIntyre, 1972). It was later found that vernalization increased the rate of flowering in the late flowering accessions (Best & Mc Intyre, 1976). Analyses of F2 progeny between the late and early flowering accessions determined the early flowering phenotype (spring annuals lacking the vernalization requirement) was determined by a single gene, with the late flowering allele being completely dominant (McIntyre & Best, 1978).

In this report we address the questions concerning the molecular basis for the spring flowering habit in pennycress, whether the spring habit has arisen only once or multiple times in nature, and we

DORN ET AL.

American Society of Plant Biologists
Cultivating a better future through plant biology research

SEB
SOCIETY FOR EXPERIMENTAL BIOLOGY

WILEY

3

begin to analyze the geographic distribution of pennycress flowering time variants. These studies were aided by genomic resources that have been developed for pennycress, including a transcriptome and draft genome (Dorn et al., 2013, 2015). These resources capture the pennycress gene space and greatly aided in the identification of alterations that lead to spring flowering. We show that key mutations in the FLC gene are responsible for spring flowering in many natural pennycress accessions, including one allele of FLC that confers the spring annual growth habit that spread from Europe to North America.

2 | MATERIALS AND METHODS

2.1 | *Thlaspi arvense* accessions and F2 population

The spring annual *Thlaspi arvense* line MN108SA was derived from a wild Minnesota population containing both winter and spring annual plants. Five generations of single seed descent was performed on a spring annual plant from this collection and sequenced. A single MN111 plant was also carried through two generations of single seed descent and sequenced.

Additional pennycress accessions described here with the "PI" or "Ames" prefix were obtained from USDA-GRIN. All pennycress accessions with the "MN" prefix were collected by Dr. Donald Wyse (University of Minnesota). The spring annual line Spring32 was obtained from Dr. Win Phippen at Western Illinois University under a Materials Transfer Agreement.

Plants were germinated on moist Berger BM2 germination mix (Berger, Inc., www.berger.ca), stratified at 4°C for 7 days, and grown in climate-controlled growth chambers at the University of Minnesota (21°C, 16 hr/8 hr day/night cycles at 100 micromoles/m^2/s PAR). The MN111 plant sequenced in this analysis was vernalized at 6 weeks post-germination at 4°C for 30 days in the dark. After vernalization, this plant was returned to the growth chamber conditions described above. The spring annual accessions were not vernalized as they flowered immediately.

2.2 | DNA isolation and Illumina genomic DNA sequencing

DNA was isolated from a single MN108SA and a single MN111 plant using the Omega Mag-Bind Plant DNA kit (Omega Bio-Tek, www.omegabiotek.com) according to the manufacturer's recommended protocol. These DNA samples were submitted to the University of Minnesota Genomics Center for sequencing on the Illumina HiSeq 2000 platform (Illumina Inc, www.illumina.com). Sequencing libraries were prepared using the Illumina TruSeq Nano DNA Sample Prep kit with an average library insert size of 460 base pairs. The Illumina Universal Adaptor and an Indexed Adaptor (MN111 – Illumina Indexed Adaptor #12 – barcode = CTTGTA, MN108SA – Illumina Indexed Adaptor #11 – barcode = GGCTAC) were used to create the sequencing libraries. Each library was sequenced on a full lane of Illumina HiSeq 2000 (100 base pair, paired-end).

DNA from the six selected European pennycress accessions was isolated using the Qiagen Plant DNeasy Mini kit (www.qiagen.com) following the manufacturer's recommended protocol. Illumina sequencing libraries were created for each European accession using the Illumina TruSeq Nano kit with the 350 bp insert protocol. These six libraries were sequenced across 1.5 lanes on the Illumina HiSeq 2500 platform (125 base pair, paired-end) with version 4 chemistry.

All raw sequencing datasets have been deposited in the NCBI Short Read Archive under BioProject PRJNA237017/SRA Accession SRP036068 (Supporting Information Table S1).

2.3 | Sequencing data quality control, read mapping

Illumina sequencing read quality and contamination were examined using FASTQC (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/). FASTQ files were filtered and trimmed to remove low quality reads and sequencing adaptors using BBDuk (https://sourceforge.net/projects/bbmap/) using the following parameters: ftl = 10 minlen = 50 qtrim = rl trimq = 10 ktrim = r k = 25 mink = 11 hdist = 1 ref = /bbmap/resources/adapters.fa tpe tbo. Sequencing reads from accessions MN111 and MN108SA had the additional BBDuk flag of ftr = 95. The resulting high quality reads from each accession were mapped to the v1 pennycress genome using Bowtie 2 and visualized in CLC Genomics Workbench.

2.4 | Analysis of FLC mutations in MN111 × MN108SA F2 population, global spring varieties, and EMS mutant

DNA was isolated from F2 progeny grown in the conditions described above using the Omega Bio-Tek Plant MagBind 96 kit according to the manufacturers recommended protocol. DNA oligos were designed to amplify the 5′ end of the pennycress FLC gene for the *flc-A* allele, approximately 100 base pairs upstream of the transcriptional start site (TaFLC_1_Forw: 5′ – CCGAGGAAGAAAAAGTAGATAGAGACA – 3′, TaFLC_1_Rev: 5′ – GAAGCTTAAAGGGGGAAAAAGGAA – 3′, Supporting Information Table S2). This amplicon was also able to identify the *flc-D* allele in MN133 and PI633414 and the *flc-α* allele from the EMS-mutagenized population as well. Polymerase Chain Reaction (PCR) was used to amplify this fragment, producing an approximately 450 bp amplicon. New England Biolabs Q5 High-Fidelity PCR Kit with 2× Master Mix was used, with the following thermal cycler conditions: (a) 98°C for 30 s, (b) 98°C for 10 s, (c) 57°C for 20 s, (d) 72°C for 20 s, (e) Go to step #2 34 times, (f) 72°C for 2 min, (g) 4°C hold. Reactions were visualized to confirm amplification of a single band using gel electrophoresis. PCR products were submitted to Beckman Coulter Genomics for PCR product purification and single pass Sanger sequencing. Amplicons were sequenced in both directions using the forward and reverse primers listed above. Sanger sequencing reads were analyzed in CLC Genomics Workbench and aligned against the pennycress MN106 reference genome at the FLC locus to identify sequence variants.

American Society of Plant Biologists

Cultivating a better future through plant biology research

SEB | SOCIETY FOR EXPERIMENTAL BIOLOGY

WILEY

2.5 | Identification of deletion alleles using PCR

A diagnostic PCR test was designed to test for the *flc-B* allele. A primer set (TaFLC_2_Forw − 5′ − ATAGTGTGCATCAACTGGTC − 3′, and TaFLC_2_Rev − 5′− CGAACCATAGTTCAGAGCTT − 3′, Supporting Information Table S2) were designed to amplify a single amplicon overlapping the 456 bp deletion (Supporting Information Figure S1). In the absence of the *flc-B* allele, a 2,088 bp fragment is produced, whereas a 1,632 bp fragment is produced in plants containing the *flc-B* allele (Figure 2b). New England Biolabs Q5 High-Fidelity PCR Kit with 2× Master Mix was used, with the following thermal cycler conditions: (a) 98°C for 30 s, (b) 98°C for 5 s, (c) 63°C for 15 s, (d) 72°C for 60 s, (e) Go to step #2 34 times, (f) 72°C for 2 min, (g) 4°C hold.

A diagnostic PCR test was developed to test for the presence or absence of the *flc-C* allele (Figure 2c, Supporting Information Figure S2). A primer set (TaFLC_4_Forw: 5′− GCGACGGTGAATAT GGAGTTGG − 3′, and TaFLC_3_Rev − 5′− GCTAATTTTTCAGC AAATCTCCCG − 3′, Supporting Information Table S2) was designed to amplify a 6,598 bp amplicon overlapping the 4,724 bp deletion. In plants containing the *flc-C* allele, this region produces a 1,874 bp amplicon, confirming the presence of this deletion. New England Biolabs Q5 High-Fidelity PCR Kit with 2× Master Mix was used, with the following thermal cycler conditions: (a) 98°C for 30 s, (b) 98°C for 5 s, (c) 67°C for 15 s, (d) 72°C for 40 s, (e) Go to step #2 34 times, (f) 72°C for 2 min, (g) 4°C hold. This amplicon was purified using the Qiagen QiaQuick PCR Purification Kit (www.qiagen.com) from accessions MN135, Ames22461, and PI650285 and subjected to Sanger sequencing to confirm the exact deletion size and coordinates. Identical 4,724 bp deletions were identified in these three accessions.

A secondary amplicon was used to confirm the absence of this deletion, with the forward primer (TaFLC_1_Forw: 5′ − CCGAGGAA GAAAAAGTAGATAGAGACA − 3′, Supporting Information Table S2) lying in the middle of the deleted region (Supporting Information Figure S2). Successful amplification using this secondary primer set (TaFLC_1_Forw and TaFLC_3_Rev − 5′ − GCTAATTTTTCAG CAAATCTCCCG − 3′, Supporting Information Table S2) produces a 1,397 bp fragment and confirms the absence of this deleted region. New England Biolabs Q5 High-Fidelity PCR Kit with 2× Master Mix was used, with the following thermal cycler conditions: (a) 98°C for 30 s, (b) 98°C for 5 s, (c) 64°C for 15 s, (d) 72°C for 40 s, (e) Go to step #2 34 times, (f) 72°C for 2 min, (g) 4°C hold.

3 | RESULTS

3.1 | Flowering time phenotypes of winter and spring annual pennycress

While identifying germplasm to be used for the pennycress breeding program, wild isolates of pennycress have been collected from various locations across North America (Supporting Information Table S3). A spring flowering pennycress variant, named MN108SA,

was isolated from a mixed population of both spring and winter annual individuals collected from Roseau, Minnesota. Single seed descent for five generations showed that the trait was stable and that vernalization was not required for floral induction. MN108SA was crossed to the vernalization-requiring accession MN111. A comparison of MN108SA and MN111 seedlings is shown in Figure 1a. MN108SA shows internode elongation soon after the formation of the first true leaves (Figure 1a insert), whereas MN111 forms a rosette. As previously reported by McIntyre & Best, 1978, the F1 derived from the cross between MN111 and MN108SA exhibited the dominant winter flowering habit, segregating in the F2 population at 38:12 winter annual:spring annual, suggesting that a single dominant gene is responsible for maintaining the winter annual phenotype.

Whole genome sequencing of the MN111 and MN108SA parent plants was used to identify key genomic differences underlying the difference in habit between winter annual and spring annual plants. DNA from each parent was sequenced on a full lane of the Illumina HiSeq 2000 platform (100 base pair, paired end). Over 334M sequencing reads were generated per parent (Supporting Information Table S1). A total of 29B high quality (post quality control) were generated per parent, representing >50× coverage of the predicted genome size of 539 Mb (Johnston et al., 2005). Trimmed and filtered reads were mapped to the winter annual MN106 draft genome sequence (Dorn et al., 2015).

3.2 | 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 noncausative effect on the spring annual growth habit (Supporting Information Figure 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 (Figure 1b, Supporting Information Figures S4 and S5). We identified a single guanine base insertion in the seventh 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 (Figure 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
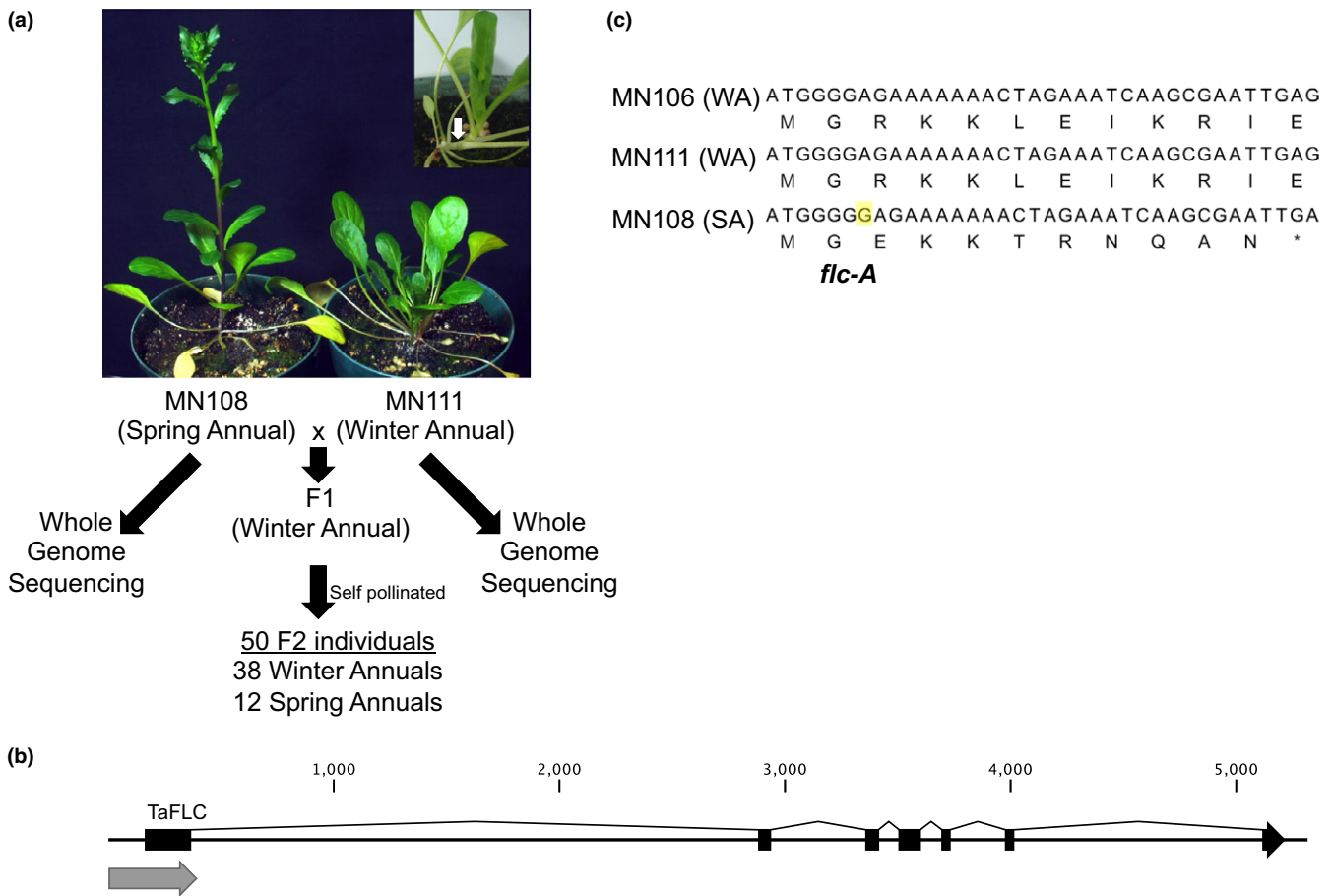
FIGURE 1 Segregation analysis confirms that *flc-A* allele confers spring annual phenotype. (a) Cogrown MN108SA and MN111 pennycress plants. Insert shows immediate internode elongation in the MN108SA spring annual plant. An F2 population was derived from a cross between MN108SA × MN111 and the segregation ratio of spring to winter annual phenotypes was determined. Whole genome sequencing of the parent plants from each background was also conducted. (b) Gene model of pennycress FLOWERING LOCUS C (FLC). Gray arrow below the first exon indicates the position of the amplicon used to genotype the SNP shown in panel C. (c) Nucleotide and predicted peptide sequence of the 5′ end of the pennycress FLC gene in winter annual MN106, winter annual MN111, and the spring annual accession MN108SA. The *flc-A* mutation in MN108SA is highlighted in yellow

exhibited a winter annual phenotype and 12 exhibited a spring annual phenotype (Figure 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, five were homozygous for the wild type FLC gene while nine were heterozygous. A chi square analysis (χ^2 = 0.026667, *p*-value = 0.8703) indicates that the data fits a 1:2:1 segregation pattern that is predicted for a single recessive mutation in this F2 population.

3.3 | Identifying a novel EMS-induced allele of FLC in pennycress

As a component of ongoing work to domesticate pennycress, an EMS-mutagenized MN106 population has been generated (Sedbrook

et al., 2014). This winter flowering population has been screened to identify early flowering variants. In the course of this screen, a new spring flowering time mutant was identified (Supporting Information Figure S6). This line did not require vernalization and flowered immediately upon germination. Whole genome sequencing of this mutant revealed a novel mutation in the first exon of FLC (*c.52C>T*), resulting in an immediate stop codon and truncated predicted protein (Supporting Information Figure S6), which was designated *flc-α*. The above segregation data along with the finding of two new independent spring flowering mutants with FLC mutations confirms the causative role of FLC mutations in the spring flowering habit in pennycress.

3.4 | Global distribution of three additional spring annual alleles of FLC

With the genetic confirmation that the *flc-A* allele was indeed the causal variant responsible for the spring annual growth habit in

MN108SA, we next examined the distribution of this allele in spring annual accessions from around the world. We screened a total of 35 spring annual accessions, which were obtained either from the University of Minnesota collections or GRIN. Of these, 23 were confirmed as pure spring annual accessions in both field and growth chamber experiments, with no winter annual segregation nor contaminating winter annual seed (Supporting Information Table S3). Surprisingly, we only identified the *flc-A* allele in three other accessions, all from North America (MN131 from Montana, USA and Ames31489 and Ames31491 from Canada).

As there were a limited number of representative spring annual accessions in our collection containing the *flc-A* allele, we utilized a PCR-based approach to determine whether there were additional mutations in FLC underlying the spring annual growth habit. As we had already amplified and sequenced the first exon of FLC in all of our spring annual accessions, we moved next to the second exon. Interestingly, we were unable to amplify the second exon in several spring annual accessions. Thus, we expanded our PCR experiments into the flanking introns around exon 2, which lead to the discovery of a second allele of FLC in spring annual lines. This mutation was characterized by a 456 bp deletion that encompasses the entire second exon of FLC, which was named *flc-B*. This allele was identified in 9 spring annual accessions distributed across the United States and Canada, along with the Spring32 accession from the Western Illinois University breeding program (Figure 2b,e, Supporting Information Table S3) (Sedbrook et al., 2014).

Having not yet identified a candidate spring annual FLC allele in any accession from Europe, we again employed whole genome sequencing of the five publicly available spring annual pennycress accessions (collected in Germany, Poland, and France). Using this approach, we were able to quickly identify a third allele, characterized by a 4,724 bp deletion (*flc-C*), which removes the first intron of FLC along with the upstream gene (gene model Ta000918), a homolog of the Arabidopsis gene UPSTREAM OF FLC (UFC, AT5G10150). In Arabidopsis, UFC and FLC are coregulated, however, UFC is not believed to influence the vernalization response (Andersson et al., 2008; Finnegan, Sheldon, Jardinaud, Peacock, & Dennis, 2004). A diagnostic amplicon experiment was designed to screen the remaining unknown spring annual accessions for the *flc-C* allele. This allele was identified in four accessions collected in Howard Springs, Montana, USA, along with three accessions from Germany, as well as accessions from Poland and France (Figure 2c,e, Supporting Information Table S3).

A fourth allele, *flc-D*, was identified through whole genome sequencing of the spring annual accession PI633414 (Germany). This allele is characterized by a single base pair change at the 100th position of the CDS and caused a precocious TAG stop codon in the predicted protein sequence (Figure 2d). As the position of this mutation was covered by the 394 bp amplicon used to identify the *flc-A* allele, we re-examined the Sanger reads from all of the accessions sequenced in our search for accessions containing the *flc-A* allele. We did not identify any additional accessions containing the *flc-D* allele.

4 | DISCUSSION

4.1 | 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.

4.2 | 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-C 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 (Figure 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 may have been introduced into North America a limited number of times, potentially from plants near Saxony or Thuringia, Germany (Figure 2e). However, it is possible that the *flc-D* allele is present in populations in North America, but was not identified in the accessions represented in this experiment.

Similarly, as we did not identify European accessions containing the *flc-A* or *flc-B* alleles, we predict that these alleles may have originated after the introduction of pennycress to North America and this allele is unique to North American accessions. It is possible that there are in fact European pennycress populations that contain the *flc-A* allele, but with the limited European germplasm availability, were not found in this study. In this scenario, it is feasible that a mixed seed lot containing *flc-A* and *flc-B* alleles were introduced to
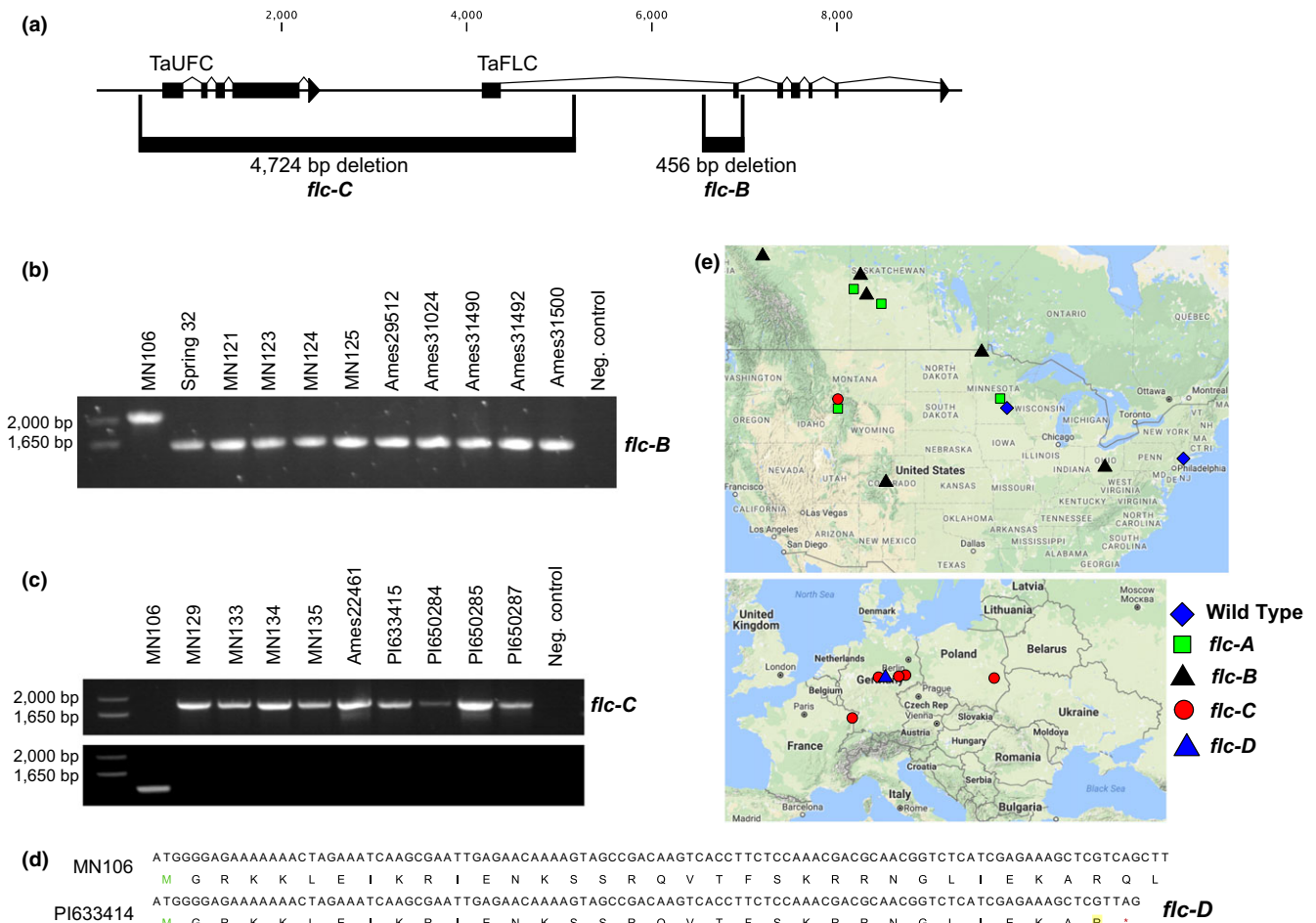
FIGURE 2 Identification of *flc-B* and *flc-C* alleles of FLC. (a) Expanded gene model of pennycress FLC region, including the upstream TaUFC gene model. Locations of the *flc-B* and *flc-C* alleles are shown below the gene model. (b) Agarose gel of the diagnostic polymerase chain reaction (PCR) results 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 (Supporting Information Figure 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 Supporting Information Table S2, and primer locations and amplicon strategy is shown in Supporting Information Figure S2. Results for accession PI650286 (Groitzsch, Saxony, Germany) are not shown, but was confirmed to have the *flc-C* allele. (d) Nucleotide and predicted peptide sequence of FLC in MN106 and PI633414 (Wachstedt, Germany), with the *flc-D* allele highlighted in yellow. (e) Global distribution of the four spring annual alleles of FLC in pennycress accessions tested in this study (listed in Supporting Information Table S3)

Montana in a single event, and plants with the *flc-A* and *flc-B* alleles spread significantly more compared to those with the *flc-D* allele. While this scenario is likely the simplest explanation, a more complex series of introduction events cannot be addressed with these data.

The accessions collected near Howard Spring, Montana (MN129, MN131, MN133, MN134, MN135) are an interesting case study of these hypotheses, especially considering the findings of the *flc-A* allele in MN131, the *flc-B* allele in MN129, MN133, MN134, and MN135, and a lack of identification of European accessions with the *flc-A* allele. The above scenarios could be investigated further in these accessions collected from Montana using genome-wide markers from either genotype-by-sequencing or WGS.

As we have already generated WGS data for five of the European accessions in this study, sequencing of these accessions from Montana and a comparison of shared polymorphisms may help in

determining the relatedness between the North American and European lines, and may point to the exact origins of these populations, as well as the corresponding FLC alleles. Again, with the limited number of pure spring annual accessions available, future studies should also focus on the collection of a more comprehensive set of globally distributed lines, including both spring and winter annuals.

4.3 | Toward an understanding of flowering time control in pennycress

Our previous efforts to develop genomic resources for the domestication of pennycress identified likely homologs controlling flowering time and the vernalization response via DNA and RNA sequence homology (Dorn et al., 2013, 2015). Here, we present the first sequence-supported genetic information on the underlying

mechanisms controlling flowering time in pennycress. While the multiple layers of evidence supporting the identified mutations in FLC as causative for the loss of vernalization requirement in spring annual lines, additional experiments are needed to eliminate the remote possibility of an unknown but tightly linked gene near the FLC loci. Additional lines of evidence to eliminate this possibility include recreating the exact spring annual mutations via CRISPR-Cas9 engineering or transgenic expression of intact (winter annual) FLC in spring annual lines of either pennycress or Arabidopsis. While the four natural alleles of FLC reported here are sufficient to confer the spring annual phenotype, there is likely a host of interacting genes and epigenetic factors also essential for the rapid flowering seen in the spring annual accessions investigated here. Of particular interest is FRIGIDA and members of the "FRIGIDA Complex" (FRI-C), including FRIGIDA LIKE 1 (FRL1), FRIGIDA-ESSENTIAL 1 (FES1), SUPPRESSOR OF FRIGIDA4 (SUF4), and FLC EXPRESSOR (FLX) (Choi et al., 2011). The FRIGIDA Complex acts as a transcriptional activation complex on FLC expression through a diverse range of functions of each complex member (Choi et al., 2011). While no loss of function FRIGIDA alleles were identified in the accessions examined here, it remains possible that an FLC-independent path to the spring annual phenotype exists in light of the limited availability of globally distributed germplasm. The interaction between the vernalization response, photoperiodic, and autonomous pathways are also of great interest, as allelic variation and unique combinations of alleles from each pathway can contribute to quantitative variation in flowering time.

As pennycress is being developed as a winter annual cover crop to fit within the fallow period in the corn—soybean rotation in the Midwestern United States, keeping a robust vernalization requirement in breeding lines is critical for pennycress to properly function as a proper winter cover crop in this system. In Minnesota, we have observed winter annual accessions that flower in the late fall or early winter after short periods of cold followed by brief warming. Occasionally, these fall-flowering plants will survive winter and continue flowering in the spring, albeit with dramatically reduced seed set. Thus, identifying the key regulators underlying a strong vernalization requirement is critical to breeding a robustly winter-annual cover crop. With an enhanced understanding of the regulators of vernalization and flowering time in pennycress, including the central role of FLC and its natural alleles identified here, we are making progress toward breeding elite cultivars with intact vernalization requirements to prevent fall flowering, yet exhibit rapid postvernalization flowering.

ACKNOWLEDGMENTS

AUTHOR CONTRIBUTIONS

KMD and MDM conceived and planned the study. EBJ, ED, and MDM constructed the F2 mapping population. EBJ and ED identified the FLC-α allele. KMD designed and performed PCR experiments, whole genome sequencing and analyses, and comparative analyses. DLW collected all "MN" designated pennycress accessions used in this study. KMD and MDM wrote the manuscript with feedback from all the coauthors.

REFERENCES

Amasino, R. M. (2005). Vernalization and flowering time. *Current Opinion in Biotechnology*, 16(2), 154–158. https://doi.org/10.1016/j.copbio.2005.02.004

Andersson, C. R., Helliwell, C. A., Bagnall, D. J., Hughes, T. P., Finnegan, E. J., Peacock, W. J., & Dennis, E. S. (2008). The FLX gene of Arabidopsis is required for FRI-dependent activation of FLC expression. *Plant and Cell Physiology*, 49(2), 191–200.

Beilstein, M. A., Al-Shehbaz, I. A., Mathews, S., & Kellogg, E. A. (2008). Brassicaceae phylogeny inferred from phytochrome A and ndhF sequence data: Tribes and trichomes revisited. *American Journal of Botany*, 95(10), 1307–1327. https://doi.org/10.3732/ajb.0800065

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*, 137, 121–127. https://doi.org/10.1086/336850

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. https://doi.org/10.1086/336670

Best, K., & McIntyre, G. (1975). The biology of Canadian weeds 9. *Thlaspi arvense* L. *Canadian Journal of Plant Science*, 55, 279–292. https://doi.org/10.4141/cjps75-039

Burn, J., Smyth, D., Peacock, W., & Dennis, E. (1993). Genes conferring late flowering in *Arabidopsis thaliana*. *Genetica*, 90(2–3), 147–155. https://doi.org/10.1007/BF01435036

Choi, K., Kim, J., Hwang, H. J., Kim, S., Park, C., Kim, S. Y., & 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. https://doi.org/10.1105/tpc.110.075911

Clarke, J. H., & Dean, C. (1994). Mapping FRI, a locus controlling flowering time and vernalization response in *Arabidopsis thaliana*. *Molecular and General Genetics*, 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. https://doi.org/10.1126/science.1221881

Crowley, J. G. (1999). *Evaluation of Camelina sativa as an alternative oilseed crop*. Carlow, Ireland: Teagasc.

Csorba, T., Questa, J. I., 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 of the United States of America*, 111(45), 16160–16165. https://doi.org/10.1073/pnas.1419030111

Dorn, K. M., Fankhauser, J. D., Wyse, D. L., & Marks, M. D. (2013). De novo assembly of the pennycress (*Thlaspi arvense*) transcriptome provides tools for the development of a winter cover crop and biodiesel feedstock. *Plant Journal*, 75(6), 1028–1038. https://doi.org/10.1111/tpj.12267

Dorn, K. M., Fankhauser, J. D., Wyse, D. L., & Marks, M. D. (2015). A draft genome of field pennycress (*Thlaspi arvense*) provides tools for the domestication of a new winter biofuel crop. *DNA Research*, 22(2), 121–131. https://doi.org/10.1093/dnares/dsu045

Elshire, R. J., Glaubitz, J. C., Sun, Q., Poland, J. A., Kawamoto, K., Buckler, E. S., & Mitchell, S. E. (2011). A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. *PLoS ONE*, 6(5), e19379. https://doi.org/10.1371/journal.pone.0019379

Evangelista, R. L., Isbell, T. A., & Cermak, S. C. (2012). Extraction of pennycress (*Thlaspi arvense* L.) seed oil by full pressing. *Industrial Crops and Products*, 37(1), 76–81. https://doi.org/10.1016/j.indcrop.2011.12.003

Fan, J., Shonnard, D. R., Kalnes, T. N., Johnsen, P. B., & Rao, S. (2013). A life cycle assessment of pennycress (*Thlaspi arvense* L.) -derived jet fuel and diesel. *Biomass and Bioenergy*, 55, 87–100. https://doi.org/10.1016/j.biombioe.2012.12.040

Finnegan, E. J., & Dennis, E. S. (2007). Vernalization-induced trimethylation of histone H3 lysine 27 at FLC is not maintained in mitotically quiescent cells. *Current Biology*, 17(22), 1978–1983. https://doi.org/10.1016/j.cub.2007.10.026

Finnegan, E. J., Sheldon, C. C., Jardinaud, F., Peacock, W. J., & Dennis, E. S. (2004). A cluster of Arabidopsis genes with a coordinate response to an environmental stimulus. *Current Biology*, 14(10), 911–916. https://doi.org/10.1016/j.cub.2004.04.045

Franzke, A., Lysak, M. A., Al-Shehbaz, I. A., Koch, M. A., & Mummenhoff, K. (2011). Cabbage family affairs: The evolutionary history of Brassicaceae. *Trends in Plant Science*, 16(2), 108–116. https://doi.org/10.1016/j.tplants.2010.11.005

Greb, T., Mylne, J. S., Crevillen, P., Geraldo, N., An, H., Gendall, A. R., & Dean, C. (2007). The PHD finger protein VRN5 functions in the epigenetic silencing of Arabidopsis FLC. *Current Biology*, 17(1), 73–78. https://doi.org/10.1016/j.cub.2006.11.052

Hojilla-Evangelista, M. P., Evangelista, R. L., Isbell, T. A., & Selling, G. W. (2013). Effects of cold-pressing and seed cooking on functional properties of protein in pennycress (*Thlaspi arvense* L.) seed and press cakes. *Industrial Crops and Products*, 45, 223–229. https://doi.org/10.1016/j.indcrop.2012.12.026

Jiang, D., Gu, X., & He, Y. (2009). Establishment of the winter-annual growth habit via FRIGIDA-mediated histone methylation at FLOWERING LOCUS C in Arabidopsis. *Plant Cell*, 21(6), 1733–1746. https://doi.org/10.1105/tpc.109.067967

Johanson, U., West, J., Lister, C., Michaels, S., Amasino, R., & Dean, C. (2000). Molecular analysis of FRIGIDA, a major determinant of natural variation in Arabidopsis flowering time. *Science*, 290(5490), 344–347. https://doi.org/10.1126/science.290.5490.344

Johnston, J. S., Pepper, A. E., Hall, A. E., Chen, Z. J., Hodnett, G., Drabek, J., … Price, H. J. (2005). Evolution of genome size in Brassicaceae. *Annals of Botany*, 95, 229–235.

Jung, C., & Muller, A. E. (2009). Flowering time control and applications in plant breeding. *Trends in Plant Science*, 14(10), 563–573. https://doi.org/10.1016/j.tplants.2009.07.005

Kim, D. H., Doyle, M. R., Sung, S., & Amasino, R. M. (2009). Vernalization: Winter and the timing of flowering in plants. *Annual Review of Cell and Developmental Biology*, 25, 277–299. https://doi.org/10.1146/annurev.cellbio.042308.113411

Marquardt, S., Raitskin, O., Wu, Z., Liu, F., Sun, Q., & Dean, C. (2014). Functional consequences of splicing of the antisense transcript COOLAIR on FLC transcription. *Molecular Cell*, 54(1), 156–165. https://doi.org/10.1016/j.molcel.2014.03.026

McIntyre, G., & Best, K. (1978). Studies on the flowering of *Thlaspi arvense* IV. Genetic and ecological differences between early and late flowering strains. *Botanical Gazette*, 139(2), 190–195.

Michaels, S. D., & Amasino, R. M. (1999). FLOWERING LOCUS C encodes a novel MADS domain protein that acts as a repressor of flowering. *Plant Cell*, 11(5), 949–956. https://doi.org/10.1105/tpc.11.5.949

Michaels, S. D., & Amasino, R. M. (2001). Loss of FLOWERING LOCUS C activity eliminates the late-flowering phenotype of FRIGIDA and autonomous pathway mutations but not responsiveness to vernalization. *Plant Cell*, 13(4), 935–941. https://doi.org/10.1105/tpc.13.4.935

Michaels, S. D., Bezerra, I. C., & Amasino, R. M. (2004). FRIGIDA-related genes are required for the winter-annual habit in Arabidopsis. *Proceedings of the National Academy of Sciences of the United States of America*, 101(9), 3281–3285. https://doi.org/10.1073/pnas.0306778101

Michaels, S. D., He, Y., Scortecci, K. C., & Amasino, R. M. (2003). Attenuation of FLOWERING LOCUS C activity as a mechanism for the evolution of summer-annual flowering behavior in Arabidopsis. *Proceedings of the National Academy of Sciences of the United States of America*, 100(17), 10102–10107. https://doi.org/10.1073/pnas.1531467100

Moser, B. R., Knothe, G., Vaughn, S. F., & Isbell, T. A. (2009). Production and evaluation of biodiesel from field pennycress (*Thlaspi arvense* L.) oil. *Energy & Fuels*, 23(8), 4149–4155. https://doi.org/10.1021/ef900337g

Phippen, W. B., & Phippen, M. E. (2012). Soybean seed yield and quality as a response to field pennycress residue. *Crop Science*, 52(6), 2767. https://doi.org/10.2135/cropsci2012.03.0192

Poland, J. A., & Rife, T. W. (2012). Genotyping-by-sequencing for plant breeding and genetics. *Plant Genome*, 5(3), 92–102. https://doi.org/10.3835/plantgenome2012.05.0005

Searle, I., He, Y., Turck, F., Vincent, C., Fornara, F., Kröber, S., … 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. https://doi.org/10.1101/gad.373506

Sedbrook, J. P., Phippen, W. B., & Marks, M. D. (2014). New approaches to facilitate rapid domestication of a wild plant to an oilseed crop: Example pennycress (*Thlaspi arvense* L.). *Plant Science*, 227, 122–132. https://doi.org/10.1016/j.plantsci.2014.07.008

Selling, G., Evangelista, R., Isbell, T., & Evangelista, M. H. (2011). Extraction of proteins from pennycress. *Abstracts of Papers of the American Chemical Society*, 241.

Shindo, C., Aranzana, M. J., 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 Physiology*, 138(2), 1163–1173. https://doi.org/10.1104/pp.105.061309

Simpson, G. G., & Dean, C. (2002). Arabidopsis, the Rosetta stone of flowering time? *Science*, 296(5566), 285–289. https://doi.org/10.1126/science.296.5566.285

Stinchcombe, J. R., Weinig, C., Ungerer, M., Olsen, K. M., Mays, C., Halldorsdottir, S. S., … Schmitt, J. (2004). A latitudinal cline in flowering time in *Arabidopsis thaliana* modulated by the flowering time gene FRIGIDA. *Proceedings of the National Academy of Sciences of the United States of America*, 101(13), 4712–4717. https://doi.org/10.1073/pnas.0306401101

Sung, S., He, Y., Eshoo, T. W., Tamada, Y., Johnson, L., Nakahigashi, K., … Amasino, R. M. (2006). Epigenetic maintenance of the vernalized state in *Arabidopsis thaliana* requires LIKE HETEROCHROMATIN PROTEIN 1. *Nature Genetics*, 38(6), 706–710. https://doi.org/10.1038/ng1795

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. https://doi.org/10.1038/nature08618

Tadege, M., Sheldon, C. C., Helliwell, C. A., Stoutjesdijk, P., Dennis, E. S., & Peacock, W. J. (2001). Control of flowering time by FLC orthologues in *Brassica napus*. *The Plant Journal*, 28(5), 545–553. https://doi.org/10.1046/j.1365-313X.2001.01182.x

Warwick, S. I., Francis, A., & Susko, D. J. (2002). The biology of Canadian weeds. 9. *Thlaspi arvense* L. (updated). *Canadian Journal of Plant Science*, 82, 803–823. https://doi.org/10.4141/P01-159

Wu, J., Wei, K., Cheng, F., Li, S., Wang, Q., Zhao, J., … Wang, X. (2012). A naturally occurring InDel variation in BraA FLC. b (BrFLC2) associated with flowering time variation in *Brassica rapa*. *BMC Plant Biology*, 12(1), 151. https://doi.org/10.1186/1471-2229-12-151

Yang, H., Howard, M., & Dean, C. (2014). Antagonistic roles for H3K36me3 and H3K27me3 in the cold-induced epigenetic switch at Arabidopsis FLC. *Current Biology*, 24(15), 1793–1797. https://doi.org/10.1016/j.cub.2014.06.047

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of the article.