PNAS
www.pnas.org

Supplementary Information for

Transposable elements drive rapid phenotypic variation in *Capsella rubella*

Xiao-Min Niu[1,2†], Yong-Chao Xu[1,2†], Zi-Wen Li[1], Yu-Tao Bian[1,2], Xing-Hui Hou[1,2], Jia-Fu Chen[1,2], Yu-Pan Zou[1,2], Juan Jiang[1,2], Qiong Wu[1], Song Ge[1,2], Sureshkumar Balasubramanian[3] and Ya-Long Guo[1,2*]

[1] State Key Laboratory of Systematic and Evolutionary Botany, Institute of Botany, Chinese Academy of Sciences, Beijing 100093, China. [2] University of Chinese Academy of Sciences, Beijing 100049, China. [3] School of Biological Sciences, Monash University, VIC 3800, Australia. [†] These authors contributed equally.

* Ya-Long Guo

Email: yalong.guo@ibcas.ac.cn

This PDF file includes:

Supplementary text

Fig. S1 to S12

Tables S1 to S8

References for SI reference citations

Supplementary Information Text

Plant material

The *C. rubella* accessions were described in our previous study (1-3) or this study (SI Appendix, Table S4). *FRI*^{sf2} *flc-3 A. thaliana* plants have been reported previously (4). Plants were grown in the greenhouse under long-day (LD) conditions (16 h light/8 h dark) at 20°C and 40–65% humidity. Flowering time was assayed as days to flowering (DTF). Flowering time was measured in the F_2 population of 879 × MTE and the BC_2F_2 population, generated from the cross between 879 and MTE with two successive backcrosses using 879 as the female and recurrent parent, and F_2 population of 844 × MTE together with their grandparents were measured in 2012 and 2016 (Fig. 2A, SI Appendix, Fig. S9A).

TE identification

TEs in the reference genomes of *C. grandiflora* (5) and *C. rubella* (5) were annotated using RepeatMasker v.4.0.6 (www.repeatmasker.org). For each genome, RepeatModeler was used for the *de novo* identification of TEs (6, 7), and LTRharvest from the Genome Tools v1.5.9 package (8) was used to identify *de novo* LTR retrotransposons. Repeats longer than 80 bp annotated by RepeatModeler were removed if they shared a sequence identity of at least 80% with homologous sequences identified in a search against the NCBI non-redundant database, as described in a previous study (9). The polymorphic TEs were identified using TEPID (10) based on the raw reads of 27 resequenced genomes of *C. rubella* accessions (11) (SI Appendix, Table S1).

To exclude the impact of incomplete genome assembly on TE enrichment, orthologous gene pairs in the two congeneric species were compared. In total, 22,205 orthologous genes were identified using InParanoid (4.1 version) (12). Genes at the start or end of each scaffold and those without a TE insertion within 2000 bp upstream of the start codon or 1000 bp downstream of the stop codon were excluded from subsequent analyses (2400 genes in *C. grandiflora*, 18 genes in *C. rubella*; 8 orthologous genes in the two species). Of the remaining 19,795 orthologous genes, the

distances from the TE insertion to the start codon (0–2000 bp, 100 bp bins) and the stop codon (0–1000 bp, 100 bp bins) of the closest gene were calculated.

Phylogenetic analysis of *C. rubella*

A neighbor-joining tree of 27 *C. rubella* accessions with resequenced genomes and the reference MTE was constructed using PHYLIP (version 3.696) (13) with 100 bootstrap replicates, based on 1,850,955 SNPs.

Transcriptome sequencing and analysis

RNA was extracted from flower bud tissues using the SV Total RNA Isolation System (Promega, Madison, WI, USA), with three biological repeats for each accession (SI Appendix, Table S2). The RNA-seq libraries were sequenced to obtain 125 paired-end reads using Illumina HiSeq 2500. Reads were mapped to the reference MTE genome using Tophat (14). Expression levels of genes were estimated by FPKM (fragments per kilobase of exon per million reads mapped) with Tophat and Cufflink softwares (15). All genes whose mean expression levels (FPKM) were larger than 3 across three accessions were retained for the following analysis (16). 607 polymorphic TE were inserted in the 2 kb upstream or downstream regions or genic regions of 345 genes across the 3 accessions with expression data (Table S2). Genes with or without TE insertions were divided into two groups, and wilcoxon rank sum tests were used to detect genes with expression levels that were significantly correlated with a TE insertion.

Mapping by sequencing

Genomic DNA was extracted from pooled leaves of early-flowering 879 × MTE (844 × MTE) F_2 plants using the CTAB (cetyltrimethyl ammonium bromide) protocol (17). In the 879 × MTE F_2 population, tissues of 104 early-flowering individuals from the 557 plants of the F_2 population that flowered earlier than 68 days after sowing were pooled and sequenced to map the causal loci. In the 844 × MTE F_2 population, pooled DNA extracted from 96 early-flowering individuals (61 days after sowing) from 501 total individuals was sequenced to map the causal loci. Briefly, 100 (150) bp paired-end reads were sequenced using the Illumina HiSeq 2000 (Illumina HiSeq X Ten) with insert

sizes of around 400 bp, and 159,506,270 (47,139,418) total reads were obtained. Short reads with about 68.9-fold (24.8-fold) coverage were mapped to the MTE reference genome, and SNPs were called using SHORE (18). The SHORE scoring matrix approach was used to detect and filter heterozygous SNPs (uniqueness of reads, coverage ≥10, minimum allele frequency ≥0.2, SNP quality score ≥25). In total, 110,091 (144,143) SNPs were retained to identify causal regions with an excess of homozygous alleles using SHOREmap (19).

DNA sequence analysis

PCR was performed to amplify the *FLC* fragment from the 5' upstream region to the 3' UTR or a partial fragment from *C. rubella* and its related species (SI Appendix, Table S5). Purified PCR products were sequenced using the ABI 3730 automated sequencer (Applied Biosystems, Foster City, CA, USA). All sequences were assembled using Lasergene SeqMan (DNASTAR, Madison, WI, USA) and aligned using Mega 5.0 (20). All primers are listed in SI Appendix, Table S8. psRNATarget (http://plantgrn.noble.org/psRNATarget/) was used to identify miRNA target sites.

ENM modeling

ENMTools 1.3 (21) was used to measure niche similarity between the two groups, as assessed by *Warren's I* similarity statistic, ranging from 0 (indicating that niches have no overlap) to 1 (niche equivalency), with 100 pseudo-replicates, as previously described (22). Ecological niche overlap and the identity test were performed using ENMtools 1.3.

Vector construction and plant transformation

For the complementation assay and the poly(A) signal mutation assay of *CrFLC* in *A. thaliana, FLC* genomic fragments (including ~2.7 kb upstream of the ATG start codon and ~1.0 kb downstream of the stop codon) were amplified using Q5 polymerase (NEB, Ipswich, MA, USA). These fragments with (or without) mutations were cloned into the pCAMBIA1300 vector using the *KpnI* and *SalI* sites and introduced into *A. thaliana* with *FRI*^{sf2} and *flc-3* alleles in the Col-0 background by *Agrobacterium tumefaciens*-mediated transformation (23). For the GUS-reporter assay, the promoter (including ~2.7

kb upstream of the ATG start codon) was amplified from MTE and cloned into the pBI121 vector using the *SbfI* and *XmaI* sites, and the 3' UTR was cloned using the *SacI* and *EcoRI* sites.

Expression analysis and 3' RACE

RNA was extracted from 14-day-old seedlings using the SV Total RNA Isolation System (Promega). First-strand cDNA was obtained using M-MLV Reverse Transcriptase (Promega). RT-qPCR (one cycle of 95°C for 20s, followed by 45 cycles of 95°C for 15s, 58°C for 45s) was performed using the ViiA^{TM} 7 Real-time PCR System (Life Technologies, Carlsbad, CA, USA) with SYBR Premix Ex TaqII (TaKaRa, Kusatsu, Japan) according to the manufacturers' instructions. The $2^{-\Delta\Delta CT}$ method was used to determine the gene expression level. Each experiment was repeated with three independent biological replicates, and RT-qPCR reactions were performed with three biological replicates and three technical replicates for each sample. All gene expression levels were normalized to the level of *β-TUBULIN*. The *FLC* 3' UTR was obtained by 3' RACE using the SMARTer RACE cDNA Amplification Kit (Clontech, Mountain View, CA, USA).

RNA stability measurement

C. rubella accessions were grown on 1/2 MS plates at 20°C. The 14-day-old seedlings were transferred to incubation buffer with 1 mM PIPES, pH 6.25, 1 mM trisodium citrate, 1 mM KCl, and 15 mM Sucrose (24). After incubation for 30 min with shaking, 3-deoxyadenosine (cordycepin; Sigma, St. Louis, MO, USA) was added to a final concentration of 200 μg/mL, and vacuum infiltration was performed for 30 s. Samples were collected at different time points to extract RNA. RT-qPCR was performed to quantify gene expression levels. *Eukaryotic initiation factor-4A* (*eiF4A*, Carubv10013828m) and *EXPANSIN-LIKE A1* (*ATEXLA1*, Carubv10017847m) were used as control genes with high and low mRNA stability, respectively (25). The expression level at each time point was first normalized to the *eiF4A* control at the corresponding time-point, and then normalized to the expression at time 0. Three biological replicates and three technical replicates for each accession were used in this experiment.

Bisulfite genomic sequencing

DNA was extracted from 20-day-old seedlings using the CTAB protocol. Bisulfite conversion was performed using the EpiTect Bisulfite Kit (Qiagen, Hilden, Germany) according to the manufacturer's instructions. The target sequences were amplified from the bisulfite DNA using Q5 High-Fidelity DNA Polymerase (NEB) and cloned using the pEASY-Blunt Simple Cloning Kit (TransGen, Beijing, China) for sequencing. For each target sequence, a minimum of 20 individual clones per sample were sequenced and analyzed. Methylation data analysis was performed with the online tool Kismeth (26).

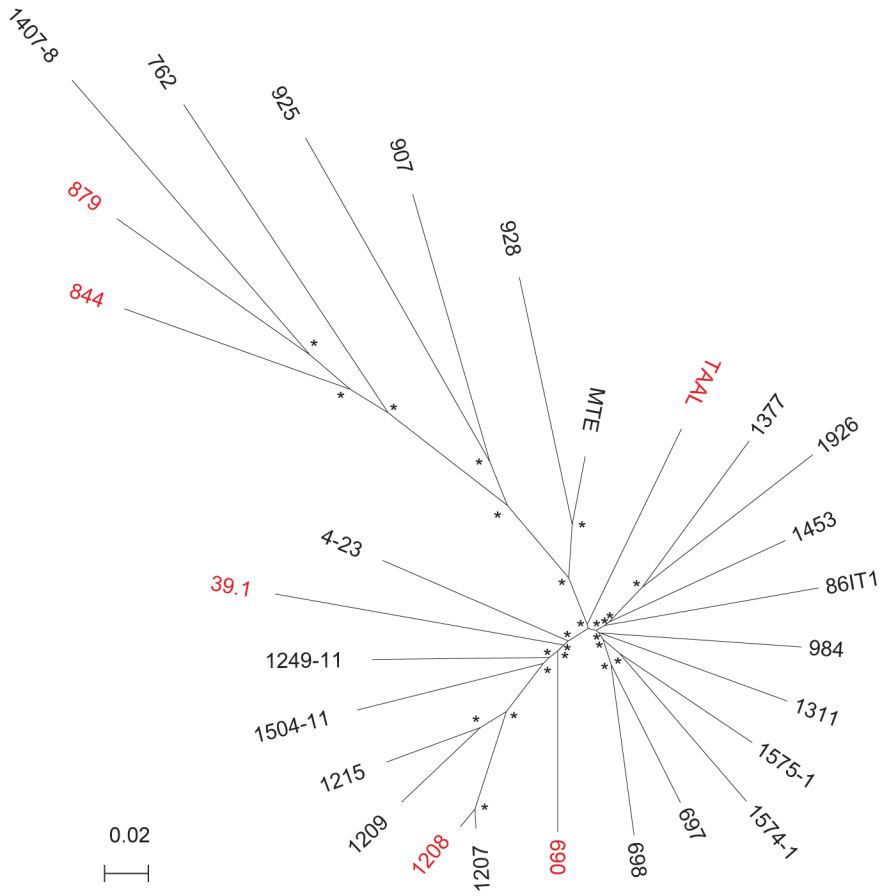
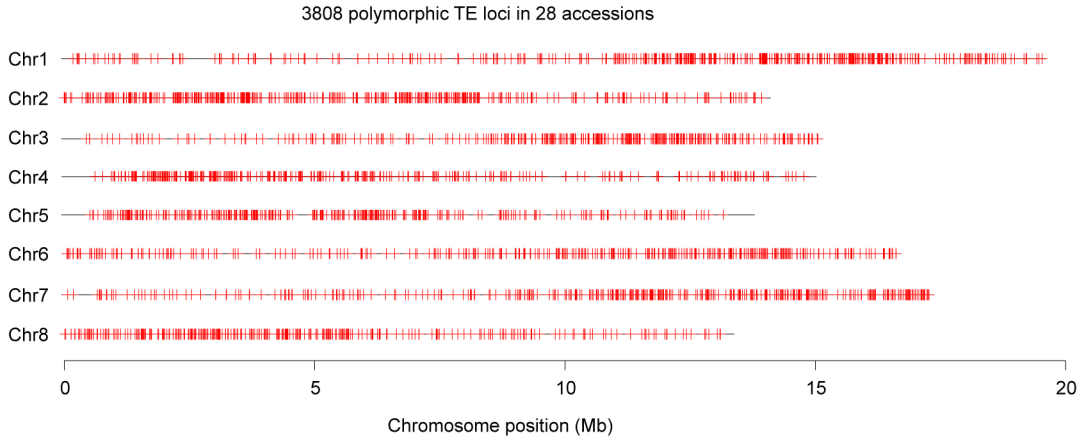


Fig. S1. Phylogenetic tree of 27 *C. rubella* accessions and the reference MTE based on whole-genome sequences. Bootstrap values larger than 95% are indicated at branches using asterisk (*). Accessions named in red indicate their *FLC* with TE insertion, in black indicate *FLC* without TE insertion.

A



B

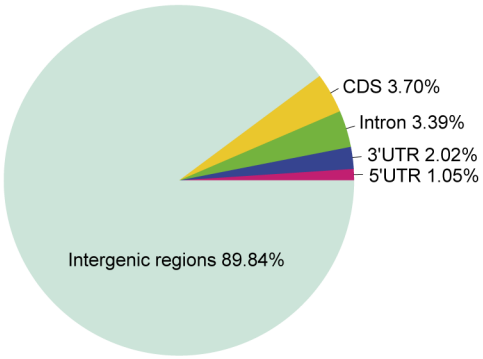


Fig. S2. Distribution of the polymorphic TEs in different genomic region of 27 *C. rubella* natural accessions and the reference MTE. A. Polymorphic TEs on chromosomes. Vertical red lines indicate TEs on the chromosomes. B. The fraction of TEs in different genic regions of the genomes.

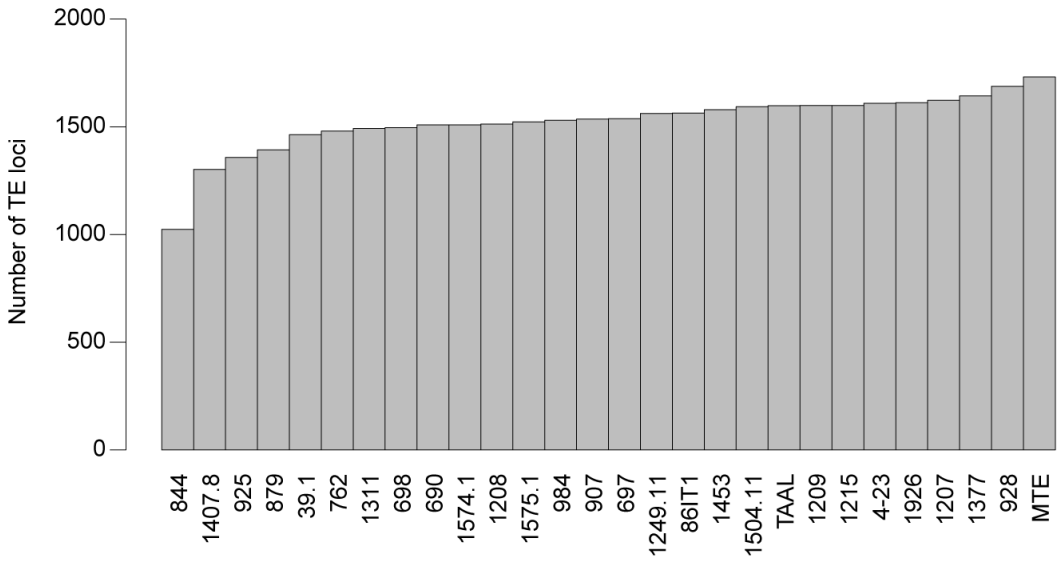


Fig. S3. The number of polymorphic TEs in the genomes of 27 *C. rubella* natural accessions and the reference MTE.

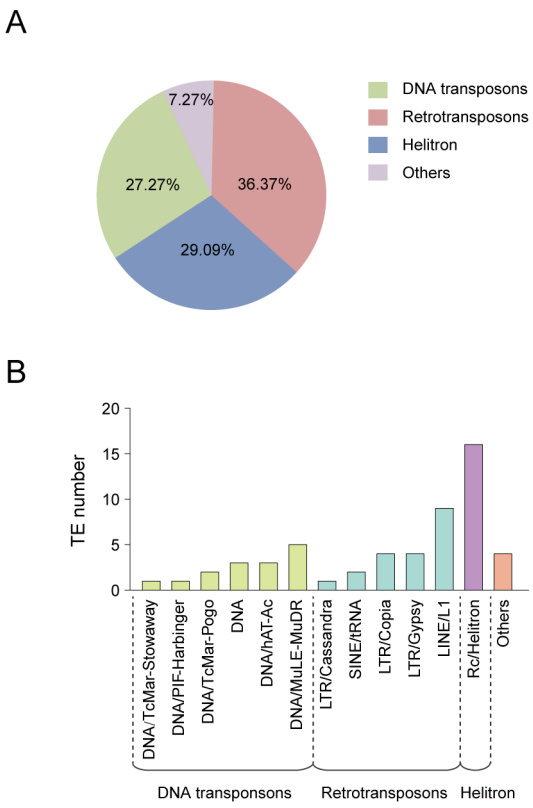


Fig. S4. Composition of the TEs that could affect gene expression levels of its adjacent genes. A. Proportion of the different TEs. B. The frequency of different TEs.

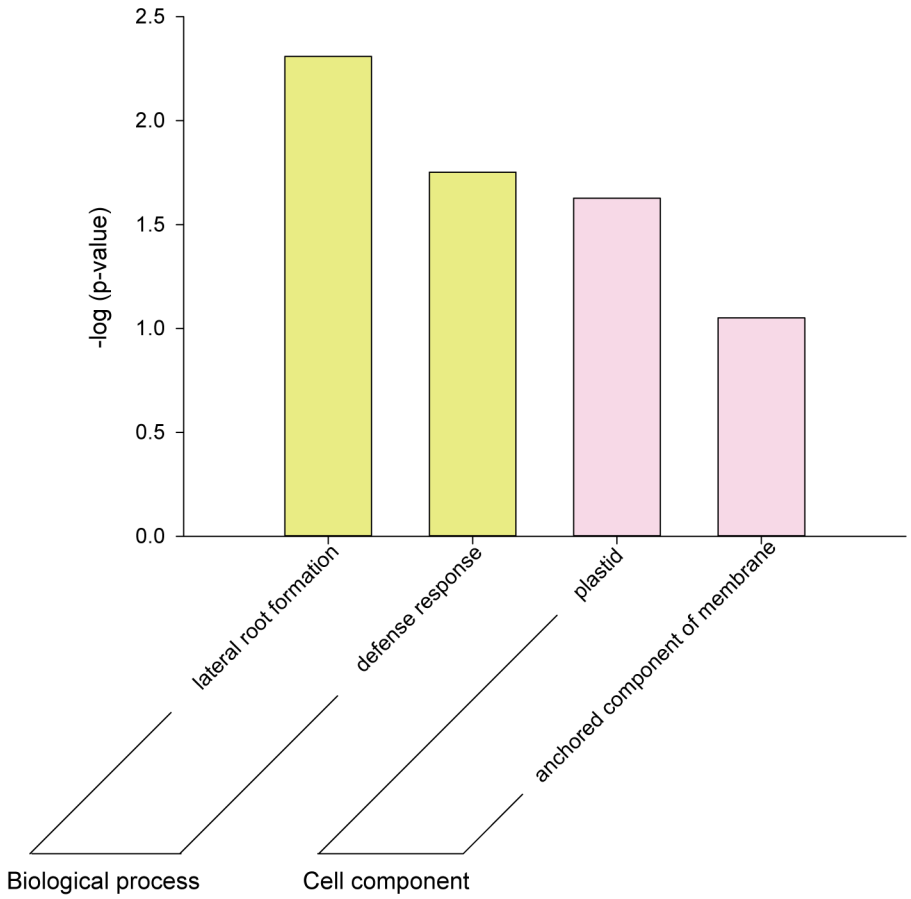
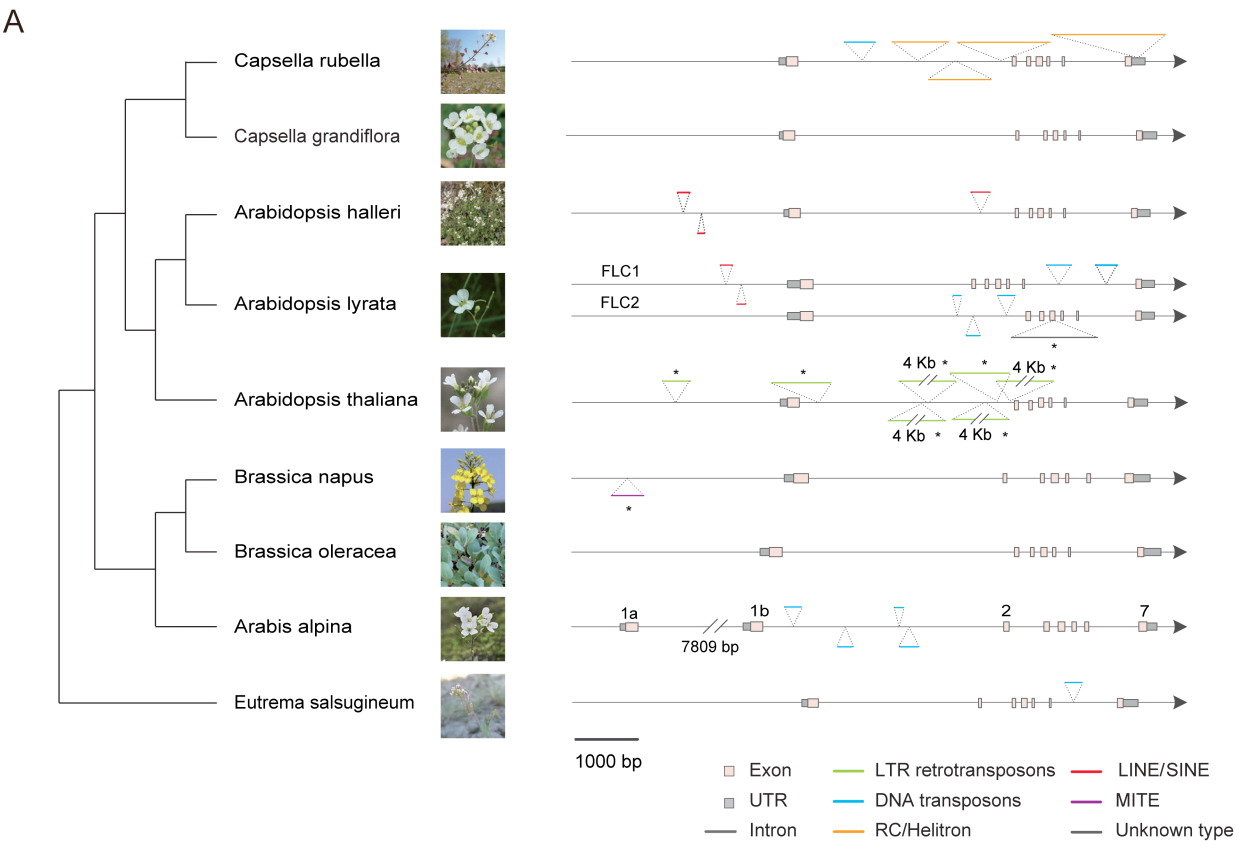


Fig. S5. GO enrichment analysis of the Arabidopsis orthologous genes of the differently expressed genes after TE insertion.



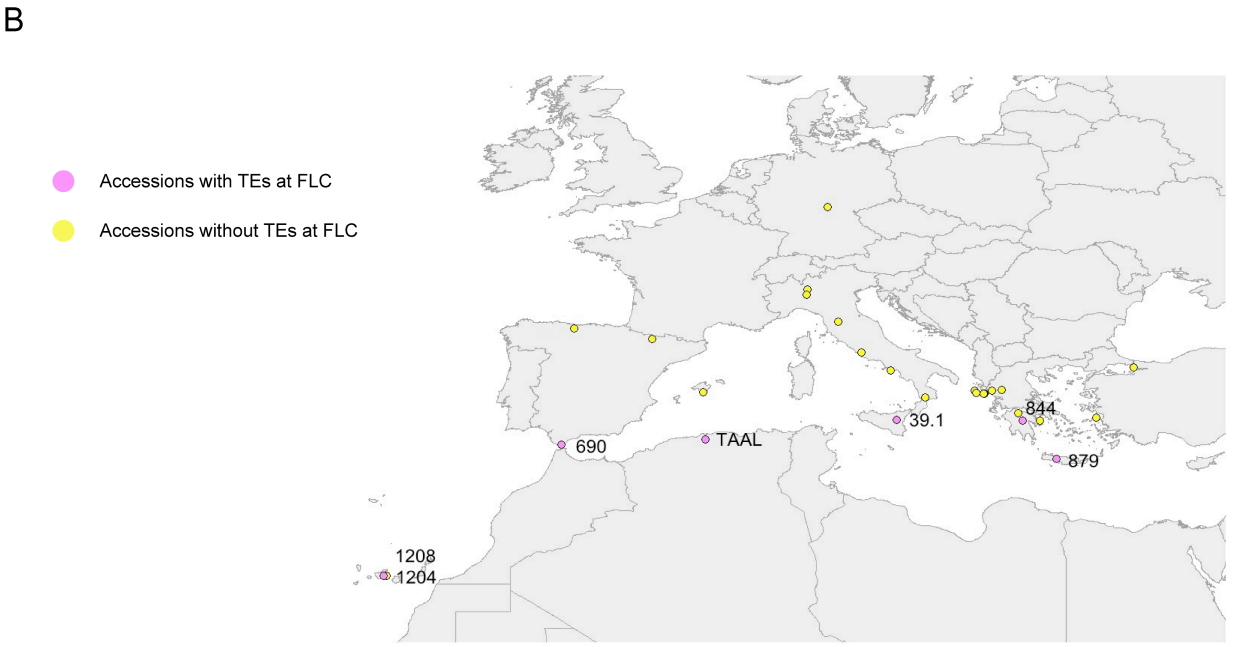


Fig. S6. Insertion of TEs on *CrFLC* gene among 35 accessions of *C. rubella* and Brassicaceae species mainly based on the orthologous sequences from reference genomes. A. Insertion of TEs on the *FLC* gene across Brassicaceae family. Dashed

lines indicate the sequence length is not scaled to real length, but marked in real value nearby. Star (*) indicates the TE insertions were confirmed in previous study, including a TE in *A. lyrata* **(27)**, and seven TEs in *A. thaliana* **(28, 29)**, and one TE in *Brassica napus* **(30)**, other TEs were annotated based on the *FLC* orthologous gene sequence extracted from the sequenced reference genome, except for *C. rubella*, of which TEs are annotated based on Sanger sequencing of natural population. The photos of plants are adapted from the websites of phytozome (https://phytozome.jgi.doe.gov/) and uniprot (http://www.uniprot.org/). B. Distribution of accessions with and without TE insertion.

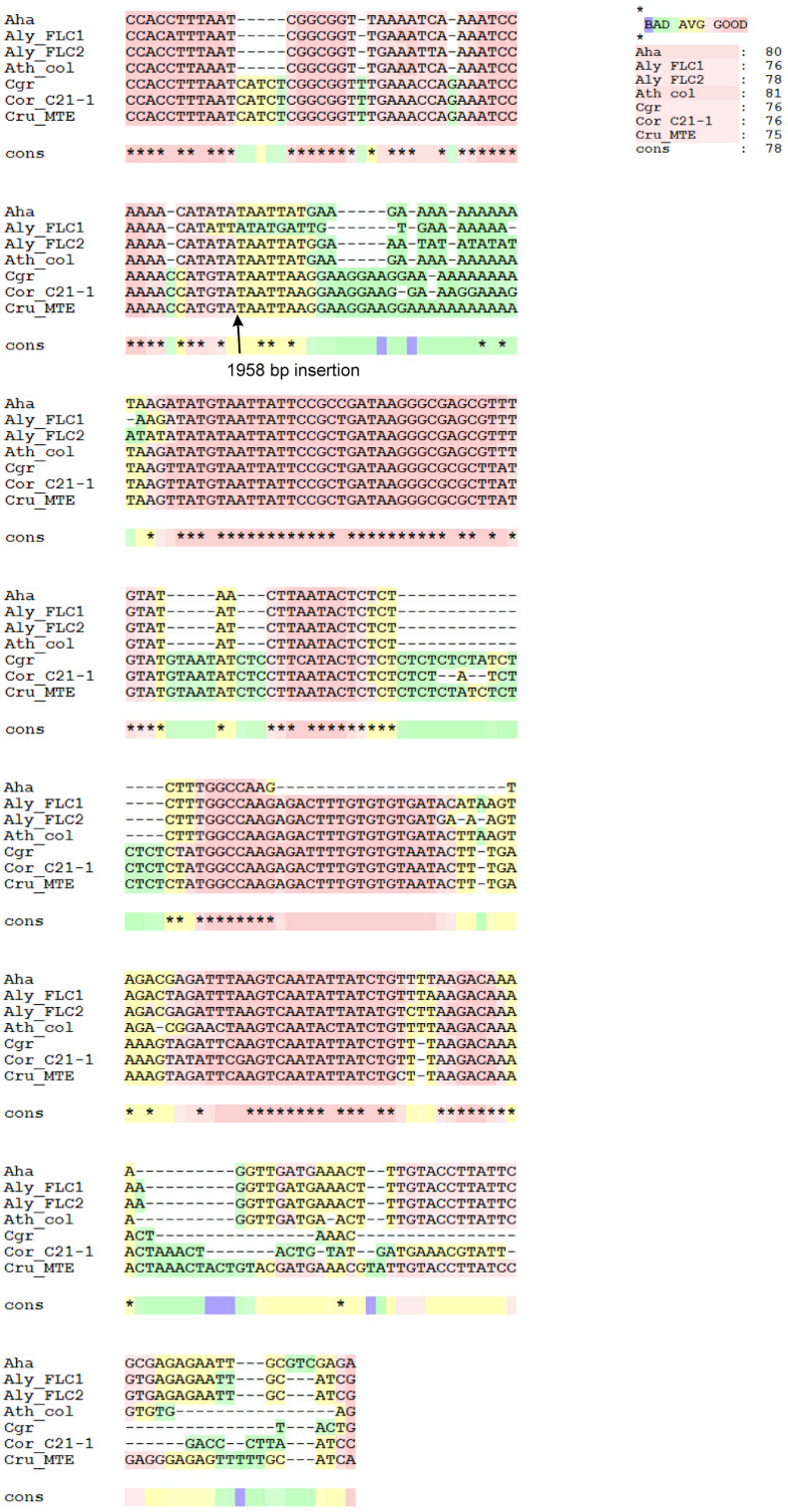
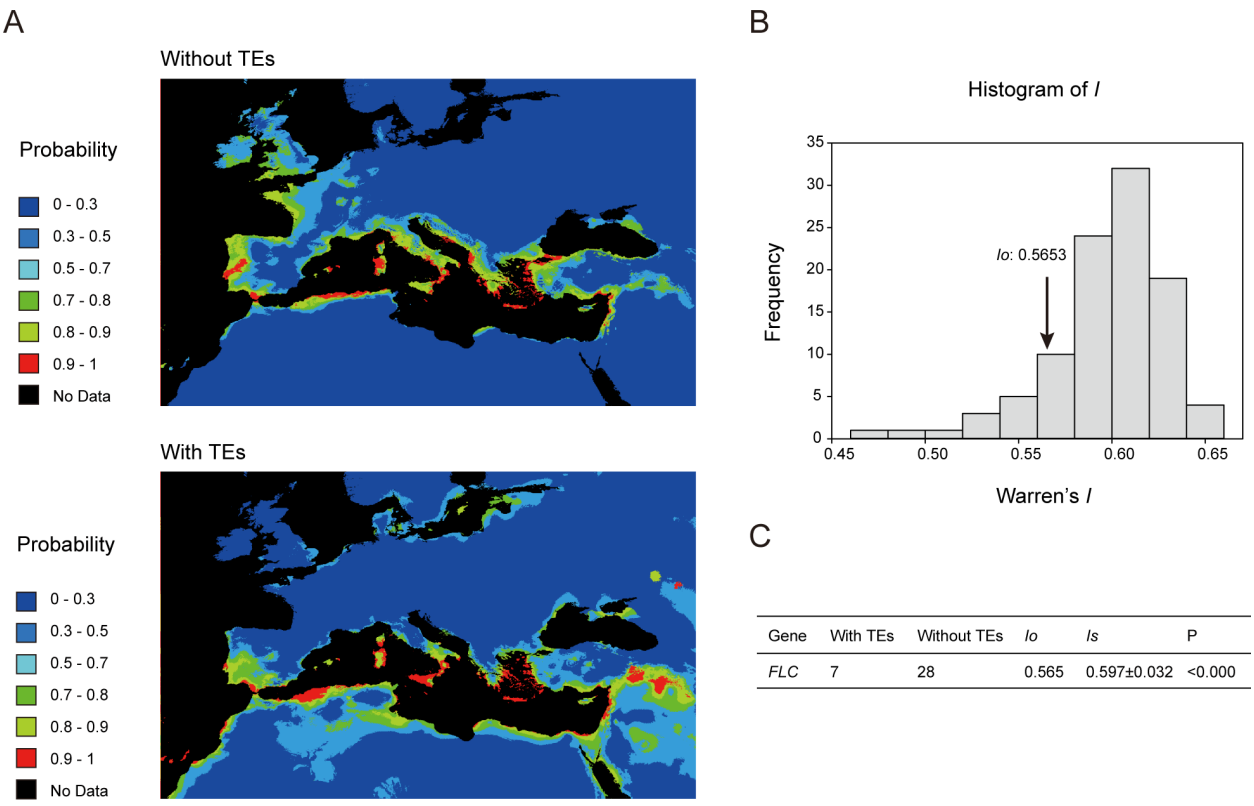


Fig. S7. Sequence variation of *FLC* 3′ UTR region among closely related species of *C. rubella*. Arrow indicates the site of the 1958-bp Helitron insertions in 879 and 1208.



Gene	With TEs	Without TEs	*Io*	*Is*	P
FLC	7	28	0.565	0.597±0.032	<0.000

Fig. S8. Ecological Niche Modeling of the Niche Differentiation between the *C. rubella* accessions with TEs and without TEs. A. Predicted distributions of *C. rubella* accessions with TEs and without TEs at the present time. B. Warren's *I* indicates the niche identity test between *C. rubella* accessions with TEs and without TEs. Arrow indicates the observed identity values, and histograms of gray bars indicate the simulated identity values (100 permutations). C. Significant ecological divergence between the two groups, as indicated by the observed *I* score (I_O) and simulated *I* scores (I_S).

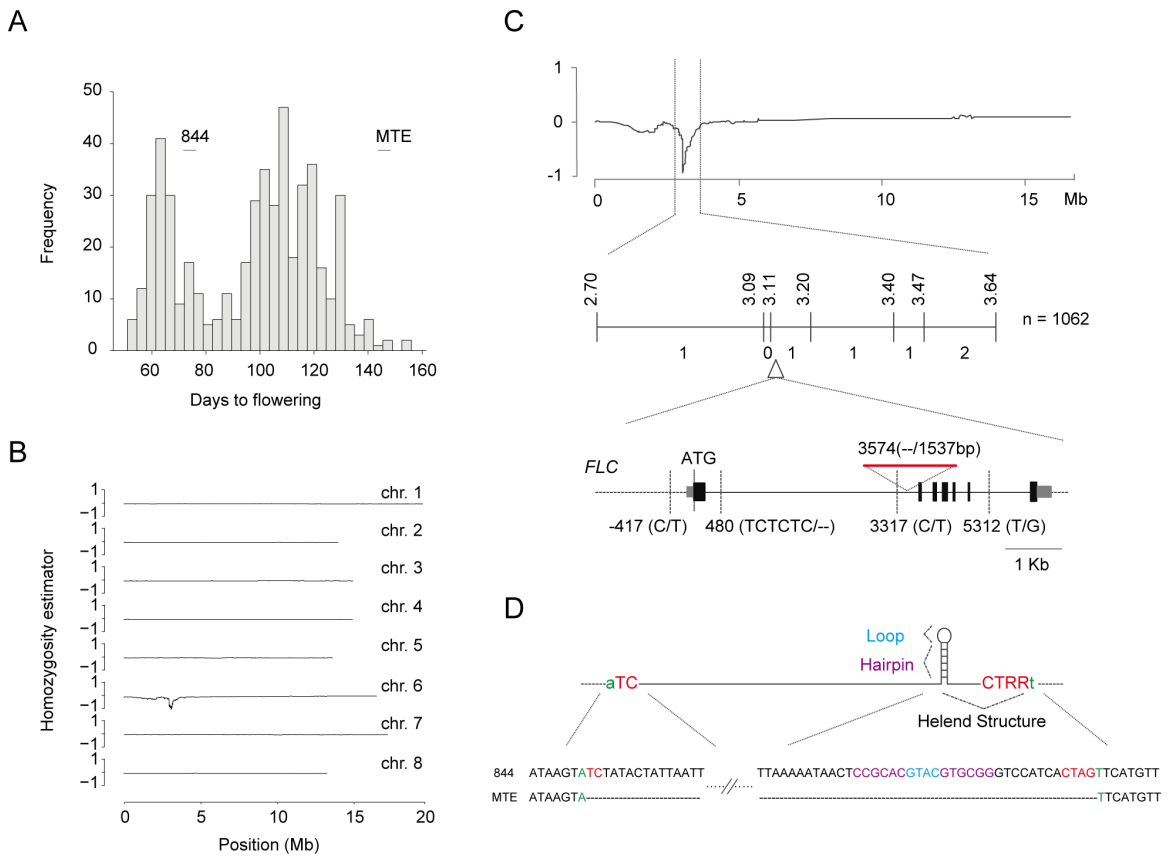


Fig. S9. QTL mapping analysis of 844 flowering time and sequence variation of the candidate gene *CrFLC*. A. Distribution of flowering time of the 501 844 × MTE F_2 individuals. Averages and ranges of flowering times for the two grandparents are shown above. B. SHOREmap analysis of flowering time. The homozygosity estimator is 0 for even allele frequencies for both parents, 1 when homozygous for the late-flowering accession MTE, and -1 when homozygous for the early-flowering accession 844. C. Fine mapping by a genetic-linkage analysis and the candidate gene *CrFLC* sequence variation. The number of recombinants between the markers and the causal locus is indicated on the top of the linkage map. TE insertion is marked in red. Dashed lines indicate sequence variation between MTE and 844. D. Sequence variation in the intron 1 and the inserted Helitron structure. The Helitron structure is illustrated at the top, which was composed of 5′ TC and 3′ CTRR termini (shown in red), two short palindromic sequences close to the 3′ terminus that could form a 20 bp hairpin (shown in blue and purple), and inserted within a host dinucleotide AT (shown in green). The

end sequences of the *CrFLC* Intron 1 Helitrons are shown at the bottom in the same color.

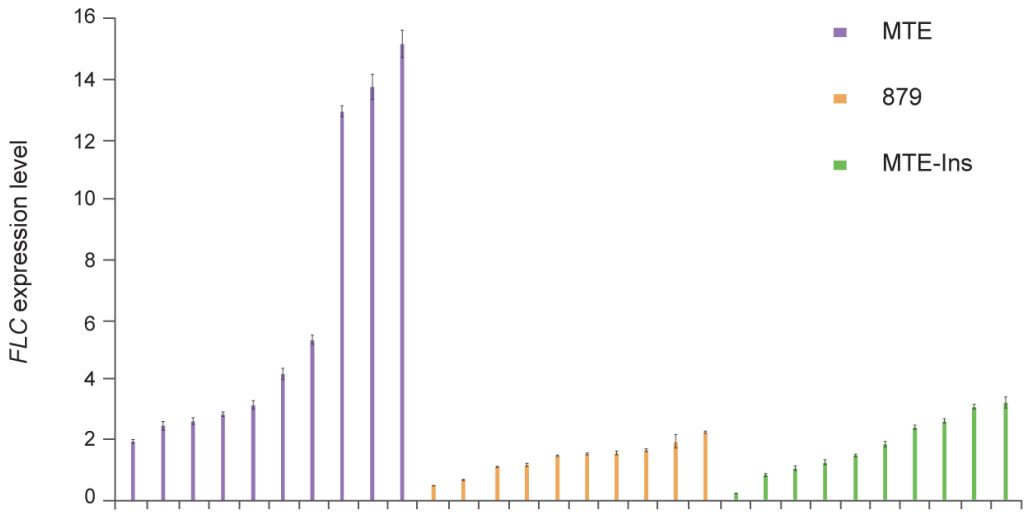
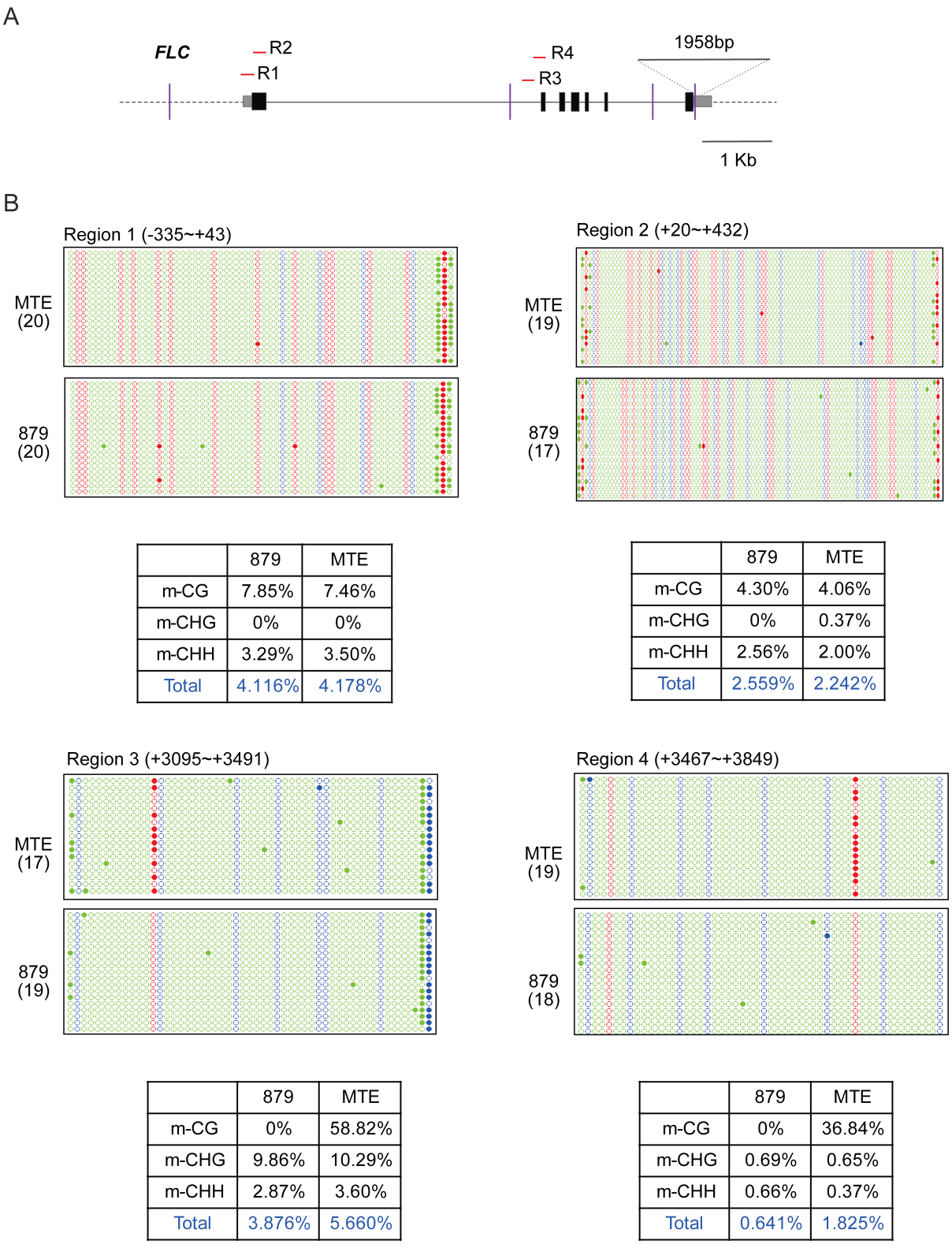


Fig. S10. Expression levels of the *CrFLC* in transgenic plants, as assayed using 10 randomly selected transgenic lines.



	879	MTE
m-CG	7.85%	7.46%
m-CHG	0%	0%
m-CHH	3.29%	3.50%
Total	4.116%	4.178%

	879	MTE
m-CG	4.30%	4.06%
m-CHG	0%	0.37%
m-CHH	2.56%	2.00%
Total	2.559%	2.242%

	879	MTE
m-CG	0%	58.82%
m-CHG	9.86%	10.29%
m-CHH	2.87%	3.60%
Total	3.876%	5.660%

	879	MTE
m-CG	0%	36.84%
m-CHG	0.69%	0.65%
m-CHH	0.66%	0.37%
Total	0.641%	1.825%

Fig. S11. DNA methylation levels between MTE and 879 at the four different regions of *FLC*. A. schematic diagram of the *C. rubella FLC* locus and of the four regions (R1-R4)

in which the levels of DNA methylation were measured. B. Percentage of CG, CHG and CHH methylation levels and dotplot analysis in MTE (top) and 879 (bottom) for the *FLC* in the four selected regions. Green, blue and red circles indicate absence (open) or presence (closed) of CHH, CHG and CG methylation, respectively.

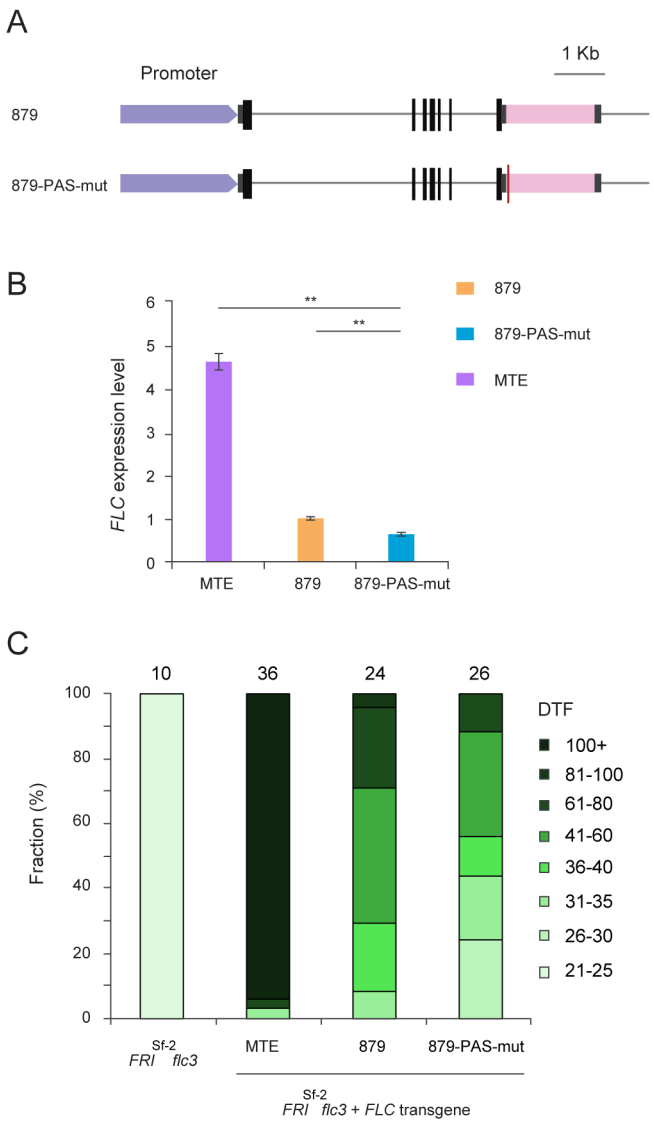


Fig. S12. Mutation the poly(A) signal in TE insertion at *CrFLC*. A. Diagrams of constructs. Red vertical line indicates the mutated poly(A) signal (mutated AAUAAA to AACAAA). Pink rectangles indicate TE insertion. B. Expression of *CrFLC* in transgenic plants, as assayed using a pool of 25 independent transgenic lines. Expression data were normalized against *AtTubulin* areshown as means ± s.d.; ** $p < 0.01$. C. Flowering time for transgenic plants. Flowering time for T1 transgenic lines with the *FRI*^{sf2} *flc-3* background are shown; the number of independent transgenic lines scored for each construct is given above the bar graph.

Table S1. *Capsella rubella* samples used in transposable elements analysis.

Accessions	Number of polymorphic TE loci	Project number in GenBank
844	1024	PRJEB6689
1407-8	1302	PRJEB6689
925	1358	PRJEB6689
879	1393	PRJEB6689
39.1	1463	PRJNA511520
762	1480	PRJEB6689
1311	1492	PRJEB6689
698	1496	PRJEB6689
690	1509	PRJEB6689
1574-1	1509	PRJEB6689
1208	1513	PRJEB6689
1575-1	1523	PRJEB6689
984	1530	PRJEB6689
907	1536	PRJEB6689
697	1538	PRJEB6689
1249-11	1562	PRJEB6689
86IT1	1564	PRJEB6689
1453	1579	PRJEB6689
1504-11	1593	PRJEB6689
TAAL	1598	PRJNA511520
1209	1599	PRJEB6689
1215	1599	PRJNA511520
4-23	1609	PRJNA511520
1926	1612	PRJNA511520
1207	1623	PRJEB6689
1377	1643	PRJEB6689
928	1688	PRJNA511520
MTE	1731	Reference genome

Table S2. Summary of RNA-Seq data.

Samples	Raw reads
879-1	41,928,270
879-2	39,686,700
879-3	42,748,052
86IT1-1	41,031,306
86IT1-2	43,497,764
86IT1-3	40,301,676
MTE-1	39,587,372
MTE-2	46,022,918
MTE-3	43,751,324

Table S3. Differently expressed genes after TEs insertion.

TE position	gene	TE location	TE present (1)/absent (0)			86IT1 expression (fpkm)			879 expression (fpkm)			MTE expression (fpkm)			P
			86IT1	879	MTE	86IT1-1	86IT1-2	86IT1-3	879-1	879-2	879-3	MTE-1	MTE-2	MTE-3	
chr1:11113888..11117858	Carubv10010764m	upstream 2kb	1	0	1	633.04	561.38	471.71	57.64	44.31	38.11	560.44	649.74	498.20	0.028
chr1:11133697..11134738	Carubv10008629m	upstream 2kb	1	0	1	84.52	103.59	89.66	67.98	68.21	63.20	80.40	81.58	78.11	0.028
chr1:11175514..11176396	Carubv10010531m	upstream 2kb	1	0	1	37.70	35.33	32.09	21.97	19.81	21.23	38.26	34.59	40.26	0.028
chr1:18549536..18550185	Carubv10011141m	upstream 2kb	1	0	1	17.13	14.41	16.65	29.86	28.68	23.76	7.63	8.99	9.74	0.028
	Carubv10012196m	upstream 2kb				87.89	79.33	82.11	97.49	95.51	93.87	72.98	74.93	70.22	0.028
chr1:18732312..18732855	Carubv10008267m	downstream 2kb	1	0	1	4.36	4.36	4.36	2.07	1.29	2.20	4.26	5.39	5.06	0.028
chr1:9161766..9161970	Carubv10008549m	downstream 2kb	0	1	0	48.78	52.35	56.39	41.08	39.83	36.86	48.53	46.08	48.99	0.028
chr1:9555103..9555603	Carubv10010343m	upstream 2kb	1	0	1	609.37	620.20	640.99	976.62	829.06	850.88	730.65	706.45	709.27	0.028
chr2:2350376..2351353	Carubv10021647m	downstream 2kb	1	0	1	20.82	24.67	22.24	44.54	48.79	56.60	18.39	23.13	24.04	0.028
chr2:2632306..2634924	Carubv10021366m	downstream 2kb	1	0	1	11.70	10.91	11.51	10.91	7.40	6.18	12.96	13.79	11.25	0.048
chr2:2875463..2876623	Carubv10020737m	upstream 2kb	1	0	1	4.96	3.82	7.35	11.59	11.42	19.16	5.73	7.09	7.12	0.028
chr2:3778860..3779211	Carubv10019783m	downstream 2kb	0	0	1	29.86	36.29	41.91	34.10	33.15	25.91	51.63	47.78	43.72	0.028
chr2:4134187..4134998	Carubv10020254m	downstream 2kb	0	0	1	16.74	22.75	21.16	25.78	22.63	25.51	29.98	25.81	27.75	0.028
chr2:6980788..6984899	Carubv10020018m	upstream 2kb	0	0	1	32.81	28.16	31.69	30.90	30.36	32.90	52.89	45.84	48.98	0.028
chr3:10575921..10575982	Carubv10015004m	upstream 2kb	1	0	0	1.32	1.41	1.41	9.32	9.66	18.10	5.85	4.25	5.08	0.028
chr3:10576466..10576535	Carubv10015004m	upstream 2kb	1	0	0	1.32	1.41	1.41	9.32	9.66	18.10	5.85	4.25	5.08	0.028
chr3:11136526..11137604	Carubv10013637m	upstream 2kb	1	0	1	3.59	3.52	2.25	10.98	8.02	8.28	4.14	3.93	2.92	0.028
chr3:1586214..1588614	Carubv10015811m	downstream 2kb	1	0	1	62.72	70.08	66.73	59.27	59.20	50.56	63.06	72.11	65.67	0.028
chr3:3516993..3517323	Carubv10015153m	upstream 2kb	0	0	1	15.57	16.27	14.64	12.48	11.40	11.92	10.52	10.94	11.64	0.048
chr3:6270241..6270289	Carubv10013209m	downstream 2kb	1	0	1	32.04	31.51	28.84	39.62	32.62	35.19	28.57	28.26	26.87	0.028
chr3:6270794..6271738	Carubv10015190m	downstream 2kb	1	0	1	1.60	2.88	2.01	6.81	6.42	4.10	3.49	6.23	2.46	0.048
chr3:7975150..7975263	Carubv10015844m	upstream 2kb	0	1	0	6.62	5.32	5.69	16.75	14.90	12.69	5.58	2.82	3.55	0.028
chr4:11346428..11346554	Carubv10023112m	CDS	1	0	0	18.78	22.46	24.21	15.59	17.47	16.95	13.68	13.76	15.21	0.028
chr4:13803494..13803710	Carubv10023542m	downstream 2kb	0	1	0	7.81	6.16	7.61	3.03	3.56	4.44	5.85	4.30	5.15	0.048
chr4:1421317..1421497	Carubv10023451m	downstream 2kb	1	0	1	5.50	4.34	5.07	3.45	3.55	2.94	5.17	4.85	4.96	0.028
chr4:1576775..1577043	Carubv10024579m	upstream 2kb	0	1	1	6.87	6.44	7.23	2.97	3.74	3.91	4.66	4.53	4.78	0.028
chr4:1737946..1739009	Carubv10023355m	upstream 2kb	0	1	1	26.86	25.99	25.58	32.40	36.68	34.29	34.22	33.96	28.74	0.028
chr4:2042299..2043330	Carubv10022898m	downstream 2kb	0	0	1	9.87	9.57	8.95	12.27	12.18	12.23	22.01	16.17	19.49	0.028
chr4:2895942..2896049	Carubv10022872m	upstream 2kb	0	1	0	24.17	23.16	22.10	60.66	55.64	61.60	22.41	18.93	20.04	0.028
chr4:2897609..2897691	Carubv10022872m	upstream 2kb	0	1	0	24.17	23.16	22.10	60.66	55.64	61.60	22.41	18.93	20.04	0.028
chr4:7244473..7244678	Carubv10024210m	downstream 2kb	0	1	1	6.89	5.98	5.91	9.45	7.88	9.91	9.45	11.46	11.71	0.028
chr4:9335301..9336229	Carubv10022696m	downstream 2kb	0	0	1	4.27	4.06	4.50	4.25	4.46	4.68	5.81	4.86	5.40	0.028
chr5:8787966..8788086	Carubv10018528m	CDS	0	1	0	0.00	0.00	0.00	10.55	7.45	12.54	0.00	0.00	0.00	0.028
chr5:9329754..9330756	Carubv10017206m	upstream 2kb	0	0	1	42.51	43.11	44.42	34.48	40.08	37.81	19.78	17.36	12.76	0.028

chr6:107886..108857	Carubv10002530m	upstream 2kb	0	0	1	5.33	5.44	6.24	5.96	5.96	4.92	4.18	3.63	1.98	0.028
chr6:13586058..13586798	Carubv10001881m	upstream 2kb	1	0	1	382.08	448.00	446.80	307.98	281.72	254.22	346.35	397.89	455.27	0.028
chr6:13621327..13621709	Carubv10002732m	upstream 2kb	0	0	1	16.26	15.71	12.93	13.45	12.98	13.11	40.87	51.75	54.80	0.028
chr6:13938478..13940601	Carubv10003418m	downstream 2kb	1	0	1	11.46	8.94	12.34	14.07	30.00	35.52	9.28	22.26	13.42	0.048
chr6:16545121..16546298	Carubv10000626m	downstream 2kb	0	0	1	22.69	20.52	20.96	26.12	25.63	27.42	16.44	17.58	15.82	0.028
chr6:1691069..1691293	Carubv10000469m	upstream 2kb	1	1	0	3.90	3.83	4.63	4.38	4.71	4.30	9.89	10.80	11.11	0.028
	Carubv10000538m	downstream 2kb				9.09	9.69	7.91	8.82	7.27	9.82	11.87	12.42	13.62	0.028
chr6:5303966..5304279	Carubv10003177m	upstream 2kb	1	0	1	19.66	18.33	20.44	0.99	1.57	2.18	11.30	6.76	13.15	0.028
chr7:10542458..10542718	Carubv10007586m	upstream 2kb	0	1	1	2.34	2.71	2.19	5.76	5.83	4.51	2.95	2.46	2.84	0.048
chr7:10870860..10871059	Carubv10004513m	downstream 2kb	0	1	0	104.90	132.80	120.50	80.37	76.37	84.02	113.92	136.41	141.33	0.028
chr7:1531935..1532293	Carubv10004997m	downstream 2kb	1	0	1	3.60	4.47	4.99	3.61	2.89	3.07	4.15	3.85	4.66	0.048
chr7:16386922..16389130	Carubv10007179m	downstream 2kb	1	0	1	12.11	13.68	16.33	31.96	32.50	17.87	23.23	13.74	16.19	0.048
chr7:17128111..17130950	Carubv10006002m	upstream 2kb	0	0	1	23.84	18.62	14.87	8.31	12.70	4.65	4.08	3.40	4.14	0.028
chr7:1980712..1981683	Carubv10005538m	upstream 2kb	1	0	1	20.66	20.35	19.55	19.24	16.57	17.25	22.19	24.34	26.42	0.028
chr7:4801769..4801930	Carubv10005339m	upstream 2kb	1	0	1	25.35	20.44	19.82	25.97	25.56	31.48	21.98	21.88	23.09	0.028
chr7:9012390..9012489	Carubv10005722m	upstream 2kb	0	0	1	10.53	8.95	9.17	9.53	9.23	10.33	7.97	8.81	8.88	0.028
chr7:9195699..9196359	Carubv10006303m	upstream 2kb	1	0	1	6.10	10.02	8.51	4.43	4.74	5.73	8.92	7.53	7.60	0.028
chr8:1069352..1069894	Carubv10025985m	upstream 2kb	1	0	1	7.82	8.50	7.73	4.81	4.58	5.84	10.85	13.65	15.55	0.028
chr8:11301054..11301170	Carubv10027847m	CDS	0	1	0	5.73	5.19	7.52	2.21	2.48	2.17	9.56	5.47	5.31	0.028
chr8:3215469..3215649	Carubv10025901m	upstream 2kb	1	0	0	6.82	7.02	6.40	10.15	10.42	10.85	15.79	15.59	16.27	0.028
chr8:4286368..4287205	Carubv10028101m	upstream 2kb	1	0	1	27.09	15.80	14.73	14.34	7.40	14.12	15.61	18.86	16.27	0.028
chr8:9098002..9098560	Carubv10025806m	upstream 2kb	1	0	1	3.78	3.41	4.05	5.90	6.19	5.58	4.21	3.97	3.88	0.028
chr8:9217013..9217341	Carubv10027175m	downstream 2kb	1	0	1	32.61	32.17	33.45	21.80	21.51	21.70	29.76	33.55	32.89	0.028

Table S4. *Capsella rubella* samples used in this study and their flowering time previously reported (3).

Accession	3'-UTR genotype	Origin (country)	Latitude/Longitude	DTF in 20LD	GenBank number or references of *FLC* genomic sequences
879	879	Greece	+35.29/+24.42	46.3	(2)
1208	879	Spain (Tenerife)	+28.19/-16.34	56.1	(2)
39.1	MTE	Bacia, Sicily, Italy	unknown	51.0	MF142149
75.2	MTE	unknown	unknown	125.2	MF142150
690	MTE	Spain	+36.15/-5.58	53.7	(2)
697	MTE	Italy	unknown	56.1	(2)
698	MTE	unknown	unknown	131.3	(2)
762	MTE	Greece	+37.58/+23.43	191.2	(2)
844	MTE	Greece	unknown	68.9	(2)
907	MTE	Greece	+39.40/+19.48	95.5	(2)
913*	MTE	unknown	unknown	43.7	MF142151
925	MTE	Greece	+39.40/+20.51	194	(2)
928	MTE	Greece	+39.46/+21.10	122.1	(3)
984	MTE	Spain	+39.30/+3.00	110.5	(3)
1204	MTE	Spain (Tenerife)	+28.19/-16.34	115.4	(2)
1207	MTE	Spain (Tenerife)	+28.19/-16.34	73.8	(2)
1209	MTE	Spain (Tenerife)	+28.19/-16.34	90.0	(2)
1215	MTE	Spain (Tenerife)	+28.19/-16.19	91.0	(2)
1297	MTE	unknown	unknown	>210	(3)
1311	MTE	France	+42.53/-0.06	171.1	(2)
1377	MTE	Argentina	+34.40/-58.30	43.9	(2)
1408	MTE	Greece	+35.29/+24.42	39.5	(2)
1411	MTE	unknown	unknown	113.0	MF142152
1482	MTE	Australia	-31,56/+115.50	>210	(2)
1856	MTE	unknown	unknown	-	MF142155
1858	MTE	unknown	unknown	191.9	(3)
1859	MTE	unknown	unknown	49.0	MF142153
1926	MTE	Italy	+45.53/+9.34	78.9	(3)
80TR1	MTE	unknown	unknown	54.1	(3)
86IT1	MTE	Sorrento, Italy	+40.62/+14.37	75.6	(3)
1GR1	MTE	Greece	+37.78/+26.83	43.2	MF142154
MTE	MTE	Italy	unknown	>210	(2)
GO665	MTE	Germany	unknown	>210	(3)
RIAH	MTE	Italy	unknown	79.3	(2)
TAAL	MTE	Algeria	unknown	57.7	MF142156

Unknown indicates no data, * indicate this accession's flowering time was measured with other accessions but not reported in our previous study.

Table S5. The sample list to compare *FLC* downstream sequence variation in two other congeners, *Capsella orientalis* and *C. grandiflora*, and relative species *A. thaliana*. Most accessions has been reported in previous studies, including all the samples of *C. grandiflora* (31) and *A. thaliana* (32), and some of *C. orientalis* has been reported in our previous study (22).

A. thaliana	***Capsella rubella***	***Capsella orientalis***			***C. grandiflora***
3'UTR	**3'UTR**	**3'UTR**	**longitude**	**latitude**	**3'UTR**
Agu-1	39.1	C12-2	46.75	90.34	924-6
Aitba-2	690	C12-3	46.75	90.34	926-8
Altenb-2	697	C12-4	46.75	90.34	927-4
Angel-1	698	C12-5	46.75	90.34	934-16
Angit-1	75.2	C12-6 #	46.75	90.34	
Apost-1	762	C12-10	46.75	90.34	
Bak-2	844	C12-11	46.75	90.34	
Bak-7	879	C13-2	46.74	90.34	
Bolin-1	907	C13-3 #	46.74	90.34	
Borsk-2	913	C13-4	46.74	90.34	
Bozen-1	925	C16-2	46.68	90.37	
Bozen-2	928	C21-1	47.21	89.78	
Castelfed-4	984	C21-2	47.21	89.78	
Castelfed-5	1204	C21-3	47.21	89.78	
Cdm-0	1207	C21-4	47.21	89.78	
Ciste-1	1208	C21-6	47.21	89.78	
Ciste-2	1209	C21-9	47.21	89.78	
Copac-1	1215	C21-11 #	47.21	89.78	
Del-10	1297	C22-7	46.73	90.35	
Dubra-1	1311	C22-9 #	46.73	90.35	
Dog-4	1377	C24-1	46.68	90.38	
Don-0	1408	C24-2	46.68	90.38	
Fy15-2	1411	C24-3 #	46.68	90.38	
Fei-0	1482	C25 #	46.7	90.37	
Galdo-1	1856	C34-3 #	47.73	88.22	
HKT2-4	1858	C43-3	45.64	82.77	
Istisu-1	1859	C43-4 #	45.64	82.77	
Jablo-1	1926	C43-5	45.64	82.77	
Kastel-1	80TRI	C43-6	45.64	82.77	
Kidr-1	86IT1	C43-7	45.64	82.77	
Kly1	IGRI	C43-8	45.64	82.77	
Kly4	MTE	C43-9	45.64	82.77	
Koch-1	G0665	C47-5 #	47.43	85.87	
Koz-2	RIAH	C52-2	48.69	87.03	

Krazo-2	TAAL	C54-2	46.85	84.19
Layo-1		C54-3 #	46.85	84.19
Lag2-2				
Leb-3				
Lecho-1				
Leo-1				
Lerik-3				
Mammo-1				
Mammo-2				
Mer-6				
Mitter				
Monte-1				
Moran-1				
Nemrat-1				
Niel-2				
Ped-0				
Petro-1				
Pra-6				
Qui-0				
Rovero-1				
Ru3.1-31				
Sha				
shigu-1				
shigu-2				
sij1				
sij2				
sij4				
slavi-1				
star-8				
stepn-1				
stepn-2				
Timpo-1				
Toufl-1				
Tu-SB30-3				
Tu-Schu-9				
Tu-v-13				
Tu-ura1-2				
Valsi-1				
Vash-1				
Vezzano-2				
Vezzano-3				
Vie-0				

# indicates the samples of *C. orientalis* that have been studied in our previous study.

Table S6. Marker analysis of the early-flowering plants in BC_2F_2 population generated from the cross between 879 and MTE with two successive backcrosses using 879 as the female and recurrent parent.

Marker	Location on Chr. 6	879	Genotype Heterozygous	MTE
P104/P105	2,780,292-2,780,931	190	4	0
P243/P244	2,984,772-2,984,883	192	2	0
P239/P240	3,050,745-3,050,898	193	1	0
P118/P119	3,088,982-3,089,508	194	0	0
P241/P242	3,113,017-3,113,120	194	0	0
P288/P289	3,185,061-3,902,892	193	1	0
P1410/P1411	3,306,624-3,306,937	193	1	0
P1412/P1413	3,438,702-3,439,240	192	2	0
P1414/P1415	3,511,739-3,512,475	191	3	0
P1416/P1417	3,643,039-3,642,405	190	4	0
P1420/P1421	3,863,733-3,863,319	187	7	0

Table S7. Marker analysis of the early flowering plants in F_2 population of 844 × MTE.

Marker	Location in Chr. 6	844	Genotype Heterozygous	MTE
P2666/P2667	2,701,873-2,702,576	260	1	0
P1942/P1943	2,877,023-2,877,651	260	1	0
P243/P244	2,984,772-2,984,883	260	1	0
P118/P119	3,088,982-3,089,508	261	0	0
P241/P242	3,113,017-3,113,120	261	0	0
P1938/P1939	3,196,389-3,197,075	260	1	0
P106/P107	3,218,793-3,219,485	260	1	0
P2690/P2691	3,396,917-3,397,173	259	2	0
P2688/P2689	3,469,489-3,470,470	258	3	0
P1416/P1417	3,642,405-3,643,039	256	5	0

Table S8. Primers used in this study.

Experiment	Name	Sequence
***FLC* sequencing**		
	P001	TTAGGGCACAAGGGGCTC
	P002	GCGGAGACGACGAGAAGAG
	P003	CGATGGTGAAGGTGAAATG
	P004	CTTTATCTGACGAGTTGTTTA
	P005	AAGGTGACTTGTCGGCTAC
	P006	TCTACTTTGTTTCTTTTCTTCC
	P007	ACTTGTCGGCTACTTTTGTTC
	P008	CGAAGCAATAGGTAGGATAGC
	P009	TCTGTAGCCTCTTCAAGTCCG
	P010	GGAGTTTTGGTCCATTTTATC
	P011	CTACATTACAACCCAGACACAC
	P012	AAGTTGTTCCCCTCCTAACC
	P013	ATAAACCATGCCGTACATTC
	P014	ATTAACATGGGCTATACGCA
	P015	AACCTTGGCAAAAAGAAAGTTAC
	P016	ACATGTATCAACCATGAAACTG
	P017	AGAAGCCGAAGAGTTCAATAGC
	P018	AAACCATTACTACTCTTGATCC
	P019	CTATTTTTTGTGTCTAATCTAC
	P020	TTTGGGGTAAACGAGAGTGATG
	P021	CCAGAAGTGCGAAAACCA
	P022	TTTGAATCATAAACTATCGA
	P023	AAACCATTACTACTCTTGAT
	P024	GAAGTGCGAAAACCACAA
	P043	ACATGGTTTTGGATTTCTGG
	P044	ACCATAGTTCAGAGCTTTTGACTG
	P045	TCTGATGCGTGCTCGATGTTG
	P046	ATGTTGAAGCTTGTTGAGAA
	P047	AAATGTTTCTTCTGCCATGC
	P048	TTCAGCAGGTTGAAAATGACA
	P049	TGACAATTGACAACCCTCCA
	P050	AGCCGGTCTTCCATTTTGTA
	P051	TCTTAAAGCCTTGGTAATACAAACA
	P052	CCAATGATCAACACTACAATGTCA
	P053	CCACTCCTTTTTATGGATTTGC
	P054	ATTCGGTCTGGTTTGAGTTGAG
	P055	AAGTTTACGGCTGTGTTTCCAT
	P056	AAGATCACCATGTTCAGCAATCA
	P057	TCTCTGGAAAGAACCTTGTCCT
	P058	CCATGTCATTAGGTTGGGGTTA
	P059	AAATTCAAACCCGTTCAATCAT
	P060	GATTTATCGTAGTTTTGTTATCCA
	P061	ATGGGGAGAAAAAAACTAGAAATCA

	P062	CTAATTAAGCAGCGGGAGAGTCAC
qRT-PCR		
a.thFLC for qRT PCR	P730	TGAGAACAAAAGTAGCCGACAA
	P731	CCGGAGGAGAAGCTGTAGA
a.thFT for qRT PCR	P732	CCCTGCTACAACTGGAACAAC
	P733	CACCCTGGTGCATACACTG
a.thSOC1 for qRT PCR	P734	ACGAGAAGCTCTCTGAAAAG
	P735	GAACAAGGTAACCCAATGAAC
a.thBETA TUBULIN for qRT PCR	P736	GAGCCTTACAACGCTACTCTGTCTGTC
	P737	ACACCAGACATAGTAGCAGAAATCAAG
crFLC qRT PCR	P065	GAGACCGCCCTTTCTGTAACTA
	P066	CAGGTGACATCTCCATCATCTC
Unspliced CrFLC	P1896	TTTTACCGCCTCTTCTGTCCC
	P1897	CCACAAGCTTGCTGCATATAATC
crBETA TUBULIN for qRT PCR	P067	AGCTTGTTGAGAATGCTGATGA
	P068	GGTCACCAAAGCTAGGGGTAGT
eIF4a	P1916	CCAGAAACTGAATGAAGTAC
	P1917	CCGAAGGCTTCTCAAAACCA
GUS for qRT PCR	P2477	ACGGGGAAACTCAGCAAGC
	P2478	TGAGCGTCGCAGAACATTACAT
crFT for qRT PCR	P980	GCCAGAACTTCAACACCC
	P981	CTTCCTCCGCAACCACTC
crSOC1 for qRT PCR	P4127	ATTCGCCAGCTCCAATAT
	P4128	ATGATCCGATGCCTTCTC
3' RACE		
	P1792	CATCCGTCGCTCTTCTCGTCGTC
	P1796	GTCTCCGCCTCCGGCAAGCTCTACA
Vector construction		
	P1755	GAAATTGGTACCCCCTTTATCTGACGAGTTG
	P1756	ATTTTTGTCGACCTAATTAAGCAGCGGGAGAGTC
	P2005	GTCAAAGCTATAAATGAATTAATAGACCACAGGATGACTTTT CTTTGTGGGAA
	P2006	TTCCCACAAAGAAAAGTCATCCTGTGGTCTATTAATTCATTT ATAGCTTTGAC
	P2007	CATTATAATATTGTCAAGTAAATTATATGATTTATATCTTTTAA TAAAGAGTTATTATTGCTATATTCTTTAC
	P2008	GTAAAGAATATAGCAATAATAACTCTTTATTAAAAGATATAAA TCATATAATTTACTTGACAATATTATAATG
	P2009	CATCTCGGCGGTTTGAAACCAGAAATCCAAAACCATG
	P2010	CATGGTTTTGGATTTCTGGTTTCAAACCGCCGAGATG
	P2156	GGAGCAAGTTTTGTAAGAAATCCTATACTTCTACAC
	P2157	GTGTAGAAGTATAGGATTTCTTACAAAACTTGCTCC
	P2170	TCCCCCCGGGGGCTTCAGTCTCCGAGAGCCCCTTG
	P2171	GCCGAGCTCCAGATGGGGAAGAATCATGTTGTTG
	P2172	CGGAATTCGATTTAAATTTAAACTGGATTAAAA
	P2173	CGGAATTCAGATGCAAAAACTCTCCCTCGGATA
	P2199	CCTATACTTCTACACTATCTACAGAGACTGCCAC

	P2200	GTGGCAGTCTCTGTAGATAGTGTAGAAGTATAG
	P2201	CCAGTTTAAATCTAAATCGAATTCCGAAGGG
	P2202	CCCTTCGGAATTCGATTTAGATTTAAACTGG
DNA methylation level		
Region1	F	TAAAGAGGTAATTAGAAAGGTAATG
	R	TTTTRTTCTCAATTCRCTTRATTT
Region2	F	AAATYAAGYGAATTGAGAAYAAAA
	R	CATCACRACATTRTTCTTCAAT
Region3	F	AATAAAAAGTATGYATATTTATGTA
	R	ACAATATTTACTRATATATRATCTT
Region4	F	AAGATYATATATYAGTAAATATTGT
	R	CCTTACATAAAACTCCTTCA
Indel		
2,780,292-2,780,931	F	TAGGGTGAGAATCCCCGACA
	R	AAACTCCTACAGTGACGGCG
2,984,772-2,984,883	F	CAAGGCGGTGATCGCATAA
	R	ATTTCCATCAATGCTATCCCC
3,050,745-3,050,898	F	CACCAAATGAAAGGGATGAC
	R	ATTCTGTCTCCTGGAACTCTAT
3,088,982-3,089,508	F	GTTAATGTATGAAATGCGACTAT
	R	AATAACCTCCGACAAAACCTT
3,113,017-3,113,120	F	AAGATTACTTTTCAGTTCCAAGC
	R	CGTTGGTGATAGTGGTGGC
3,185,061-3,902,892	F	GATCGACGGTGGAGATCACA
	R	TGTCTCACCTGCTTCACTGTC
3,306,624-3,306,937	F	AGGCTGAACCGCTTGTCG
	R	CGGAAGGTGAGCCAGTAA
3,438,702-3,439,240	F	GGAATCGGGAAATGTAGC
	R	CGGGACGACTCAGGTGAA
3,511,739-3,512,475	F	CAACAACGAGCTGGTTTC
	R	TAATGGTGATTTGGAGGC
3,643,039-3,642,405	F	CAACAAGTTCAAGGCACG
	R	TCAGTACGGTAAGCACCA
3,863,733-3,863,319	F	GTCTTCGGAGCAACAAAC
	R	TGGGCATGATTAGGTGAA

References

1. Guo YL*, et al.* (2009) Recent speciation of *Capsella rubella* from *Capsella grandiflora*, associated with loss of self-incompatibility and an extreme bottleneck. *Proc Natl Acad Sci U S A* 106(13):5246-5251.

2. Guo YL, Todesco M, Hagmann J, Das S, & Weigel D (2012) Independent *FLC* mutations as causes of flowering-time variation in *Arabidopsis thaliana* and *Capsella rubella*. *Genetics* 192(2):729-739.

3. Yang L*, et al.* (2018) Parallel evolution of common allelic variants confers flowering diversity in *Capsella rubella*. *Plant Cell* 30(6):1322-1336.

4. Michaels SD & Amasino RM (1999) *FLOWERING LOCUS C* encodes a novel MADS domain protein that acts as a repressor of flowering. *Plant Cell* 11(5):949-956.

5. Slotte T*, et al.* (2013) The *Capsella rubella* genome and the genomic consequences of rapid mating system evolution. *Nat Genet* 45(7):831-835.

6. Bao Z & Eddy SR (2002) Automated de novo identification of repeat sequence families in sequenced genomes. *Genome Res* 12(8):1269-1276.

7. Price AL, Jones NC, & Pevzner PA (2005) De novo identification of repeat families in large genomes. *Bioinformatics* 21 Suppl 1(suppl_1):i351-358.

8. Ellinghaus D, Kurtz S, & Willhoeft U (2008) LTRharvest, an efficient and flexible software for de novo detection of LTR retrotransposons. *BMC Bioinformatics* 9:18.

9. Hu TT*, et al.* (2011) The *Arabidopsis lyrata* genome sequence and the basis of rapid genome size change. *Nat Genet* 43(5):476-481.

10. Stuart T*, et al.* (2016) Population scale mapping of transposable element diversity reveals links to gene regulation and epigenomic variation. *Elife* 5:e20777.

11. Ågren JA*, et al.* (2014) Mating system shifts and transposable element evolution in the plant genus *Capsella*. *BMC Genomics* 15(1):602.

12. Berglund AC, Sjolund E, Ostlund G, & Sonnhammer EL (2008) InParanoid 6: eukaryotic ortholog clusters with inparalogs. *Nucleic Acids Res* 36(Database issue):D263-266.

13. Feisenstein J (1989) PHYLIP-phylogeny inference package (version 3.2). *Cladistics* 5:164-166.

14. Trapnell C, Pachter L, & Salzberg SL (2009) TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics* 25(9):1105-1111.

15. Trapnell C*, et al.* (2012) Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. *Nat Protoc* 7(3):562-578.

16. Meng D*, et al.* (2016) Limited contribution of DNA methylation variation to expression regulation in *Arabidopsis thaliana*. *PLoS Genet* 12(7):e1006141.

17. Doyle JJ & Doyle JL (1987) A rapid DNA isolation procedure from small quantities of fresh leaf tissues. *Phytochem Bull* 19:11-15.

18. Ossowski S*, et al.* (2008) Sequencing of natural strains of *Arabidopsis thaliana* with short reads. *Genome Res* 18(12):2024-2033.

19. Schneeberger K*, et al.* (2009) SHOREmap: simultaneous mapping and mutation identification by deep sequencing. *Nat Methods* 6(8):550-551.

20. Tamura K*, et al.* (2011) MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol Biol Evol* 28(10):2731-2739.

21. Warren DL, Glor RE, & Turelli M (2010) ENMTools: a toolbox for comparative studies of environmental niche models. *Ecography* 33(3):607-611.

22. Han TS*, et al.* (2015) Frequent introgressions from diploid species contribute to the adaptation of the tetraploid Shepherd's purse (*Capsella bursa-pastoris*). *Mol Plant* 8(3):427-438.

23. Weigel D & Glazebrook J (2002) *Arabidopsis: a laboratory manual.* (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY).

24. Seeley KA, Byrne DH, & Colbert JT (1992) Red light-independent instability of oat phytochrome messenger-rna invivo. *Plant Cell* 4(1):29-38.

25. Golisz A, Sikorski PJ, Kruszka K, & Kufel J (2013) *Arabidopsis thaliana* LSM proteins function in mRNA splicing and degradation. *Nucleic Acids Res* 41(12):6232-6249.

26. Gruntman E*, et al.* (2008) Kismeth: analyzer of plant methylation states through bisulfite sequencing. *BMC Bioinformatics* 9:371.

27. Kemi U*, et al.* (2013) Role of vernalization and of duplicated *FLOWERING LOCUS C* in the perennial *Arabidopsis lyrata*. *New Phytol* 197(1):323-335.

28. Lempe J*, et al.* (2005) Diversity of flowering responses in wild *Arabidopsis thaliana* strains. *PLoS Genet* 1(1):109-118.

29. Zhai J*, et al.* (2008) Small RNA-directed epigenetic natural variation in *Arabidopsis thaliana*. *PLoS Genet* 4(4):e1000056.

30. Hou J*, et al.* (2012) A Tourist-like MITE insertion in the upstream region of the *BnFLC.A10* gene is associated with vernalization requirement in rapeseed (*Brassica napus* L.). *BMC Plant Biol* 12:238.

31. Paetsch M, Mayland-Quellhorst S, & Neuffer B (2006) Evolution of the self-incompatibility system in the Brassicaceae: identification of S-locus receptor kinase (*SRK*) in self-incompatible *Capsella grandiflora*. *Heredity* 97(4):283-290.

32. Cao J*, et al.* (2011) Whole-genome sequencing of multiple *Arabidopsis thaliana* populations. *Nat Genet* 43(10):956-963.