Transposable elements drive rapid phenotypic variation in Capsella rubella

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

aState Key Laboratory of Systematic and Evolutionary Botany, Institute of Botany, Chinese Academy of Sciences, 100093 Beijing, China; bUniversity of Chinese Academy of Sciences, 100049 Beijing, China; and cSchool of Biological Sciences, Monash University, VIC 3800, Australia

Edited by Ian T. Baldwin, Max Planck Institute for Chemical Ecology, Jena, Germany, and approved February 19, 2019 (received for review July 4, 2018)

Rapid phenotypic changes in traits of adaptive significance are crucial for organisms to thrive in changing environments. How such phenotypic variation is achieved rapidly, despite limited genetic variation in species that experience a genetic bottleneck is unknown. Capsella rubella, an annual and inbreeding forb (Brassicaceae), is a great system for studying this basic question. Its distribution is wider than those of its congenerspecies, despite an extreme genetic bottleneck event that severely diminished its genetic variation. Here, we demonstrate that transposable elements (TEs) are an important source of genetic variation that could account for its high phenotypic diversity. TEs are (i) highly enriched in C. rubella compared with its outcrossing sister species Capsella grandiflora, and (ii) 4.2% of polymorphic TEs in C. rubella are associated with variation in the expression levels of their adjacent genes. Furthermore, we show that frequent TE insertions at FLOWERING LOCUS C (FLC) in natural populations of C. rubella could explain 12.5% of the natural variation in flowering time, a key life history trait correlated with fitness and adaptation. In particular, we show that a recent TE insertion at the 3′ UTR of FLC affects mRNA stability, which results in reducing its steady-state expression levels, to promote the onset of flowering. Our results highlight that TE insertions can drive rapid phenotypic variation, which could potentially help with adaptation to changing environments in a species with limited standing genetic variation.

Brassicaceae | Capsella rubella | natural variation | rapid phenotypic variation | transposable elements

Author contributions: Y.-L.G. designed research; X.-M.N., Y.-C.X., Z.-W.L., Y.-T.B., X.-H.H., J.-F.C., and J.J. performed research; Y.-L.G. contributed new reagents/analytic tools; X.-M.N., Y.-C.X., Z.-W.L., Y.-P.Z., and Q.W. analyzed data; and X.-M.N., S.G., S.B., and Y.-L.G. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

This open access article is distributed under Creative Commons Attribution-NonCommercial-NoDerivatives License 4.0 (CC BY-NC-ND).

Data deposition: The data described in the paper have been deposited in the National Center for Biotechnology Information Sequence Read Archive under accession nos. PRJNA392709, PRJNA392711, and PRJNA511520 and in GenBank under accession nos. MF422379-MF422413, MF422414-MF422449, MF422460-MF422525, MF422489-MF422494, and MG492014-MG492020.

Significance

The mechanisms underlying rapid adaptation to changing environments in species with reduced genetic variation, referred to as the “genetic paradox of invasion,” are unknown. We report that transposable elements (TEs) are highly enriched in the gene promoter regions of Capsella rubella compared with its outcrossing sister species Capsella grandiflora. We show that a number of polymorphic TEs in C. rubella are associated with changes in gene expression. Frequent TE insertions at FLOWERING LOCUS C of C. rubella affect flowering-time variation, an important life history trait correlated with fitness. These results indicate that TE insertions drive rapid phenotypic variation, which could potentially help adapting to novel environments in species with limited genetic variation.

Author contributions: Y.-L.G. designed research; X.-M.N., Y.-C.X., Z.-W.L., Y.-T.B., X.-H.H., J.-F.C., and J.J. performed research; Y.-L.G. contributed new reagents/analytic tools; X.-M.N., Y.-C.X., Z.-W.L., Y.-P.Z., and Q.W. analyzed data; and X.-M.N., S.G., S.B., and Y.-L.G. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

This open access article is distributed under Creative Commons Attribution-NonCommercial-NoDerivatives License 4.0 (CC BY-NC-ND).

Data deposition: The data described in the paper have been deposited in the National Center for Biotechnology Information Sequence Read Archive under accession nos. PRJNA392709, PRJNA392711, and PRJNA511520 and in GenBank under accession nos. MF422379-MF422413, MF422414-MF422449, MF422460-MF422525, MF422489-MF422494, and MG492014-MG492020.

Significance

The mechanisms underlying rapid adaptation to changing environments in species with reduced genetic variation, referred to as the “genetic paradox of invasion,” are unknown. We report that transposable elements (TEs) are highly enriched in the gene promoter regions of Capsella rubella compared with its outcrossing sister species Capsella grandiflora. We show that a number of polymorphic TEs in C. rubella are associated with changes in gene expression. Frequent TE insertions at FLOWERING LOCUS C of C. rubella affect flowering-time variation, an important life history trait correlated with fitness. These results indicate that TE insertions drive rapid phenotypic variation, which could potentially help adapting to novel environments in species with limited genetic variation.

Author contributions: Y.-L.G. designed research; X.-M.N., Y.-C.X., Z.-W.L., Y.-T.B., X.-H.H., J.-F.C., and J.J. performed research; Y.-L.G. contributed new reagents/analytic tools; X.-M.N., Y.-C.X., Z.-W.L., Y.-P.Z., and Q.W. analyzed data; and X.-M.N., S.G., S.B., and Y.-L.G. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

This open access article is distributed under Creative Commons Attribution-NonCommercial-NoDerivatives License 4.0 (CC BY-NC-ND).

Data deposition: The data described in the paper have been deposited in the National Center for Biotechnology Information Sequence Read Archive under accession nos. PRJNA392709, PRJNA392711, and PRJNA511520 and in GenBank under accession nos. MF422379-MF422413, MF422414-MF422449, MF422460-MF422525, MF422489-MF422494, and MG492014-MG492020.

Significance

The mechanisms underlying rapid adaptation to changing environments in species with reduced genetic variation, referred to as the “genetic paradox of invasion,” are unknown. We report that transposable elements (TEs) are highly enriched in the gene promoter regions of Capsella rubella compared with its outcrossing sister species Capsella grandiflora. We show that a number of polymorphic TEs in C. rubella are associated with changes in gene expression. Frequent TE insertions at FLOWERING LOCUS C of C. rubella affect flowering-time variation, an important life history trait correlated with fitness. These results indicate that TE insertions drive rapid phenotypic variation, which could potentially help adapting to novel environments in species with limited genetic variation.

Author contributions: Y.-L.G. designed research; X.-M.N., Y.-C.X., Z.-W.L., Y.-T.B., X.-H.H., J.-F.C., and J.J. performed research; Y.-L.G. contributed new reagents/analytic tools; X.-M.N., Y.-C.X., Z.-W.L., Y.-P.Z., and Q.W. analyzed data; and X.-M.N., S.G., S.B., and Y.-L.G. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

This open access article is distributed under Creative Commons Attribution-NonCommercial-NoDerivatives License 4.0 (CC BY-NC-ND).

Data deposition: The data described in the paper have been deposited in the National Center for Biotechnology Information Sequence Read Archive under accession nos. PRJNA392709, PRJNA392711, and PRJNA511520 and in GenBank under accession nos. MF422379-MF422413, MF422414-MF422449, MF422460-MF422525, MF422489-MF422494, and MG492014-MG492020.
environments in plants and could explain the genetic paradox of invasion.

Results

TEs Are Highly Enriched in *C. rubella* and Affect the Expression Levels of Its Adjacent Genes. We hypothesized that if TEs are among the key determinants of phenotypic variation, their distribution may differ between *C. rubella* and its outcrossing congeneric sister species *C. grandiflora*. To test this, we compared the distribution of TEs in 19,795 orthologous gene pairs between *C. rubella* and *C. grandiflora*. TEs are highly enriched in different genic regions of the *C. rubella* compared with the corresponding regions of *C. grandiflora* (Fig. 1A). Our analysis with orthologous regions unmasked these divergent distribution patterns compared with the previous study (18), since the genomes of these two species have varied coverages/assemblies. In addition, TEs are highly enriched in the promoter and downstream regions of genes in *C. rubella* compared with *C. grandiflora* (Fig. 1B), indicating that TEs could potentially contribute to the diversification of gene expression in *C. rubella*.

To assess whether TEs are polymorphic in *C. rubella*, we scanned 28 *C. rubella* genomes, which included 27 accessions (18, 19) (*SI Appendix, Table S1*), with MTE accession as a reference (20). A phylogenetic tree based on whole-genome sequences suggested that these 28 *C. rubella* accessions are closely related (*SI Appendix, Fig. S1*). We identified 3,808 polymorphic TEs among these genomes (*SI Appendix, Fig. S2A*). As reported earlier (18), these polymorphic TEs are enriched (89.84%) in intergenic regions (*SI Appendix, Table S2*). As reported earlier, we scanned 28 *C. rubella* genomes, which included 27 accessions (18, 19) (*SI Appendix, Table S1*), with MTE accession as a reference (20). A phylogenetic tree based on whole-genome sequences suggested that these 28 *C. rubella* accessions are closely related (*SI Appendix, Fig. S1*). We identified 3,808 polymorphic TEs among these genomes (*SI Appendix, Fig. S2A*). As reported earlier (18), these polymorphic TEs are enriched (89.84%) in intergenic regions (*SI Appendix, Table S2*). Among these TEs that are associated with changes in gene expression, retrotransposons are more common (36.37%, *SI Appendix, Fig. S4*). A Gene Ontology enrichment analysis (21) of the genes that are affected by the TEs using the *Arabidopsis* orthologous genes suggested that these genes are involved in lateral root formation and defense response (*SI Appendix, Fig. S5*). Taken together, these findings suggest that polymorphic TEs can diversify gene expression, which has the potential to shape the phenotypic variation in *C. rubella* as in other plants (8, 22, 23).

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{fig1.png}
\caption{Transposable elements cause phenotypic diversification and changes in gene expression in *C. rubella*. (A) Number of TEs in different genic regions of the *C. rubella* and *C. grandiflora* genomes. (B) Enrichment of TEs in 100-bp bins in the regions 2 kb upstream of the start codon and 1 kb downstream of the stop codon between orthologous gene pairs of *C. rubella* and *C. grandiflora*. The x axis indicates the distance from TEs to the start codon (stop codon) of the closest gene; the y axis indicates the number of genes with TEs at different distances to the start codon (stop codon). (C) Frequency distribution of polymorphic TEs in the genomes of 27 *C. rubella* accessions with MTE as a reference. The x axis indicates the number of accesses with the same TE insertions. (D) Insertion of TEs in FLC in *C. rubella* populations. Thick lines in different colors and lengths indicate the classification and length of the inserted TEs. Numbers in parentheses indicate the sequence lengths of inserted TEs; the names before the parentheses indicate the accesses with the inserted TEs; when there are two accesses, accession names are separated by a slash (/). (E) Flowering-time variation between accesses with or without a TE insertion at the CrFLC locus. (F) Relative expression levels of CrFLC, CrFT, and CrSOC1 in accesses with TE insertions. Four biological replicates, each with three technical replicates, were analyzed. Expression data were normalized with CrTubulin and then to the levels of 879 plants and are shown as means ± SD; *P < 0.05; **P < 0.01; NS, insignificant.}
\end{figure}
Frequent TE Insertions at CrFLC Affect Natural Variation in Flowering Time. The geographic distribution of *C. grandiflora* is much narrower than that of *C. rubella*, which has colonized most areas of southern and western Europe with varied environmental conditions (24). Flowering time is an important trait for adaptation to changing climates and novel environments (25, 26). Interestingly, we found that *FT*, a determinant of natural variation in flowering time of *C. rubella* (27), has TE insertions in its genomic region (27). To assess the frequency of TE insertions at CrFLC, we sequenced the CrFLC genomic region (from 2 kb upstream of the start codon to the end of 3′ UTR) in eight accessions by Sanger sequencing and compared them with 27 other previously published CrFLC genomic sequences (SI Appendix, Table S4). We found five different TE insertions in 7 of 35 accessions (20%). We found a TE insertion in the 3′ UTR in two accessions (879 and 1208) and four additional TE insertions in intron 1 in five accessions (391, 690, 1204, 844, and TAAL) (Fig. 1D). The TE insertions ranged from 498 to 1,958 bp, and most were Helitron, except one insertion in the TAAL accession, which was a DNA transposon. All these seven accessions with TE insertions are early flowering compared with the reference accession MTE and accessions without TEs (Fig. 1E and SI Appendix, Table S4). The expression levels of CrFLC are lower (P < 0.05, Student’s t test) than MTE in all of the other six accessions except for 1204 (Fig. 1F). This is also reflected in the inversely correlated expression levels of the CrFLC-downstream genes *FT* and *SOC1* (Fig. 1F). Furthermore, 12.5% of the variation in flowering time could be explained by allelic variation of the TE insertions at CrFLC in *C. rubella* (Pearson’s correlation coefficient, r² = 0.125, P < 0.05). These results revealed that TE insertions in CrFLC are associated with early flowering in *C. rubella*.

Given that each TE insertion is present in only a few *C. rubella* accessions at different regions of the FLC locus, we concluded that these TE insertions originated recently and independently (SI Appendix, Fig. S1). In the Brassicaceae family, there are frequent TE insertions at FLC (SI Appendix, Fig. S6A). An analysis of a few accessions from *C. grandiflora* and *C. orientalis* and 76 accessions of *Arabidopsis thaliana* (SI Appendix, Table S5) also suggested that the 3′ UTR TE insertion in accession 879 is a recent mutation that is present only in two *C. rubella* accessions (SI Appendix, Fig. S7). If the TE insertions have any potential to be of adaptive significance, one might speculate that these accessions could show some geographic patterns. To assess this, we compared the origin of these accessions. Intriguingly, accessions with TE insertions clustered in the southernmost habitat of this species around the Mediterranean region (SI Appendix, Fig. S6B and Table S4). This region harbors a Mediterranean climate, which is characterized by high precipitation, warm winters, and dry and hot summers (28-31). Plants in the area are usually early flowering to avoid dry and hot summers (28-31). The high TE insertion frequency, associated with early flowering in *C. rubella*, is consistent with this hypothesis. To assess whether there is significant niche divergence between *C. rubella* accessions with or without the TE insertion, we performed ecological niche modeling and found significant niche divergence between the two groups (100 random permutations, one-sample t test, P < 0.001, SI Appendix, Fig. S8), consistent with the possibility that TEs in *C. rubella* show the potential to cause phenotypic variation of adaptive significance.

To test whether TE insertions in the 3′ UTR and intron 1 are indeed the causal mutations for early flowering, we performed genetic mapping. First, we analyzed an F2 population (879 × MTE) derived from a cross between the late-flowering, sequenced accession MTE (20) and the early flowering accession 879 that harbors a TE in the 3′ UTR. Flowering time correlates significantly with FLC expression in the Brassicaceae family (32–34). Here, we found that the distribution of flowering time is consistent with the distribution of *CrFLC* expression in F2 populations (Fig. 2A and B). The F2 individuals with lower *CrFLC* expression belong to the 879 genotype (Fig. 2B). Furthermore, the CrFLC expression level was strongly correlated with the TE insertion in the 3′ UTR in 2-wk-old seedlings of the F2 population (Pearson’s correlation coefficient, r² = 0.462, P < 0.01). These findings suggested that the 3′ UTR TE potentially affected *CrFLC* expression in this cross.

Sequencing of a pool of early flowering plants from the F2 population mapped the causal loci to an ∼683-kb region on chromosome 6 (from position 2.78–3.64 Mb) (Fig. 2C). We further used 825 individuals from the BC1F2 population and narrowed the causal region down to a 130-kb interval (SI Appendix, Table S6). This mapping interval contained 39 ORFs, including *CrFLC* (Fig. 2D). Sanger sequencing revealed that *CrFLC* has one deletion of 2 bp in the promoter, two SNPs in its intron, one SNP, and a 1,958-bp Helitron TE insertion in the 3′ UTR that differentiated MTE from 879 (Fig. 2D and E). The CrFLC expression level in 879 was a quarter of that in MTE (Fig. 1F), making it a prime candidate for the observed flowering-time variation.

To assess whether the TE insertion in intron 1 of CrFLC is also the causal mutation for early flowering, we constructed an F2 population of 844 × MTE. Using a similar sequencing approach, we identified an ∼650-kb region on chromosome 6 based on 501 F2 individuals (SI Appendix, Fig. S9 A and B). We narrowed down the mapping interval to a 210-kb region using 561 more F2 individuals and additional markers, which included the CrFLC locus (SI Appendix, Fig. S9 C and D and Table S7). The location of TEs in intron 1 was roughly similar to the position of *Ler* in *A. thaliana*, in which a TE insertion regulates histone modification via siRNA, thereby regulating flowering time (35). Therefore, it is likely that a similar mechanism may mediate the onset of early flowering in 844.

However, the TE insertion at the 3′ UTR in accession 879 is unique (SI Appendix, Fig. S6). To confirm whether the TE insertion in the 3′ UTR modulates flowering time, we generated constructs with genomic CrFLC from MTE and 879 and an additional construct of MTE-CrFLC in which the Helitron TE from 879 was introduced (MTE-Ins) and compared their ability to complement *FRI*′Δ2 *flc-3 A. thaliana* plants (Fig. 3A). CrFLC expression was higher in transgenic plants harboring the MTE construct than in those with 879 or MTE-Ins constructs (Fig. 3B; see SI Appendix, Fig. S10, for 10 randomly selected independent transgenic lines). Consistent with these findings, levels of the CrFLC downstream genes *CFI* and *CrSOC1* were inversely correlated with the CrFLC expression level (Fig. 3B). Flowering time was correlated with CrFLC expression. Transgenic plants with the MTE genomic sequence were late flowering, while transgenic plants with either 879 or MTE with the TE insertion were earlier flowering (all P < 0.001, Student’s t test; Fig. 3C). Taken together, these results suggested that the allelic variation caused by the TE insertion at the 3′ UTR of CrFLC was responsible for the flowering-time difference between 879 and MTE accessions of *C. rubella*.

**TE in 3′ UTR of CrFLC Causes Alternative Polyadenylation and Affects Its mRNA Stability.** TE insertions can change gene expression by altering transcription and/or by modulating mRNA stability (8). To assess whether the differential CrFLC expression between MTE and 879 is caused by transcriptional differences, we quantified transcription rates by measuring unspliced mRNA abundance (18, 36). The abundance of unspliced *CrFLC* nascent transcript is similar between MTE and 879 (Fig. 4A). Consistent with this, in both accessions, low levels of DNA methylation were detected around the CrFLC transcription start site and the end of intron 1, two regions important for the regulation of FLC expression (37) (SI Appendix, Fig. S11). These findings suggest that transcriptional differences do not account for the observed changes in gene expression.

8690 | www.pnas.org/cgi/doi/10.1073/pnas.1811498116

Niu et al.
UTR can give rise to alternative polyc- 

cle-3 A. thaliana alleles in the mRNA (36).

SD; ** and are shown as means

UTR and the inserted Helitron structure. The Helitron structure is A. thaliana = 3

SOC1 expression in the construct in TE insertion at (Fig. 4 FRI UTR causes APA, other mechanisms also contribute to the reduction in gene expression.

To assess mRNA stability, we measured CrFLC expression in the presence of cordycepin, which inhibits transcription (40), in parental lines. These experiments revealed that 879-CrFLC mRNA degraded faster than MTE-CrFLC mRNA (Fig. 4C). To further confirm the effect of the 3′ UTR, we generated GUS-reporter constructs in FRI′3 fle-3 A. thaliana plants driven by the CrFLC promoter from MTE with either MTE or 879 3′ UTRs (Fig. 4D).

We found that the MTE 3′ UTR led to an almost threefold increase in expression compared with that of 879. These results are consistent with findings from Arabidopsis, indicating that the 3′ UTR could affect the stability of FLC mRNA (36).

MTE and 879 also differ in AU-rich elements (AREs) and GU-rich elements (GREs) before their poly(A) tails (Fig. 4B). These elements, which are present only in 879, have the potential to affect RNA stability (41). To assess whether AREs or GREs could mediate changes in gene expression, we compared expression levels of GUS after specifically mutagenizing these elements in transgenic lines (Fig. 4D). While the loss of the first ARE had a significant impact on gene expression, the loss of the GRE and the second ARE did not, suggesting that the first ARE at the 3′ UTR of 879 is critical for modulating mRNA expression.

TE insertions at the 3′ UTR can give rise to alternative polyadenylation (APA), which in turn could modulate mRNA stability (38, 39). The Helitron insertion in 879 introduces a canonical poly(A) signal (AAUAAA). The 3′ RACE assays revealed that there is APA in 879-CrFLC (Fig. 4B). To assess whether APA could fully account for the observed phenotypic differences, we generated an additional 879-CrFLC construct in FRI′3 fle-3 A. thaliana plants, in which the poly(A) signal was mutated (879-PAS-mut) (SI Appendix, Fig. S12). These plants were also early flowering like transgenic plants of 879-CrFLC compared with transgenic plants of MTE-CrFLC, which suggests that, while the TE insertion at the 3′ UTR causes APA, other mechanisms also contribute to the reduction in gene expression.

Fig. 2. QTL mapping analysis of flowering time. (A) Distribution of flowering time of the 557 individuals of the 879 × MTE F2 population. Averages and ranges of flowering times for the two parental lines are shown above. (B) Histogram of CrFLC expression levels variations among F2 population (n = 122) with TE insertion (879 genotype, shown in yellow), without TE insertion (MTE genotype, purple), and heterozygote (gray) at its 3′ UTR. Averages and ranges of CrFLC expression levels for the two parental lines are shown above. Expression data were normalized against CrTubulin. (C) 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 879. (D) Fine mapping by a genetic-linkage analysis and sequence variation in the candidate gene CrFLC. The number of recombinants between the markers and the causal locus is indicated below the linkage map. CrFLC is marked in red. Dashed red lines indicate sequence variation between MTE and 879; red line indicates deleted sequence. (E) Sequence variation in the CrFLC 3′ UTR and the inserted Helitron structure. The Helitron structure is illustrated at the top, composed of the 5′ UTR, we generated GUS-reporter constructs in MTE with either MTE or 879 3′ UTRs (Fig. 4D). We found that the MTE 3′ UTR led to an almost threefold increase in expression compared with that of 879. These results are consistent with findings from Arabidopsis, indicating that the 3′ UTR could affect the stability of FLC mRNA (36).

MTE and 879 also differ in AU-rich elements (AREs) and GU-rich elements (GREs) before their poly(A) tails (Fig. 4B). These elements, which are present only in 879, have the potential to affect RNA stability (41). To assess whether AREs or GREs could mediate changes in gene expression, we compared expression levels of GUS after specifically mutagenizing these elements in transgenic lines (Fig. 4D). While the loss of the first ARE had a significant impact on gene expression, the loss of the GRE and the second ARE did not, suggesting that the first ARE at the 3′ UTR of 879 is critical for modulating mRNA expression.

Fig. 3. TE insertion at CrFLC explains variation in flowering time. (A) Diagrams of constructs. Red vertical lines indicate variation between MTE and 879. Pink rectangles indicate the TE insertion. (B) Expression of CrFLC and two downstream flowering regulators (FT, SOC1) in A. thaliana transgenic plants, as assayed using a pool of 25 independent transgenic lines. Expression data were normalized against AtTubulin and then to the levels of transgenic plants with 879-CrFLC and are shown as means ± SD; **P < 0.01, (C) Flowering time for transgenic plants. Flowering times for T1 transgenic lines with FRI′3 and fle-3 alleles in the A. thaliana Col-0 background are shown; the number of independent transgenic lines scored for each construct is given above the bar graph. DTF, days to flowering.

Niu et al.
in these transgenic lines. An analysis of transgenic lines harboring reporter constructs with partial UTRs from MTE suggested that the latter part of the MTE 3′ UTR is more critical as the deletion of the first half of the MTE 3′ UTR (MTEΔ2) resulted in the same expression level as that of MTE, but the deletion of the latter half (MTEΔ1) significantly reduced the expression level (Fig. 4D). These results indicate that the presence of AREs and the absence of the latter part of the MTE 3′ UTR in 879-CrFLC might reduce CrFLC expression probably by affecting mRNA stability (Fig. 4D).

Discussion

Invasive species with limited genetic variation often rapidly acquire diversification of phenotypes to adapt to changing environments and become successful in a process referred to as the genetic paradox of invasion (2). How this process occurs is a fundamental question in biology. C. rubella experienced an extreme genetic bottleneck but has a wide distribution and is therefore a good model system to address this question. TEs have been suggested as an agent of rapid adaptation because they can produce abundant genetic variation in a limited time (5–9) and can affect phenotypic variation (11–13). For example, in peppered moths, a TE insertion within intron 1 of the cortex gene increases its expression level and induces melanization (13). However, the extent to which TEs contribute to rapid phenotypic diversification and the mechanisms by which they influence phenotypes are unknown.

Our results show that TEs play a crucial role in rapid phenotypic variation and suggest that they can potentially aid in adaptation to changing environments. Previous studies in other systems have showed that TEs in 3′ UTRs can alter gene expression via the regulation of methylation in Arabidopsis (42) and can change the translation efficiency in rice (43). We have found that a TE insertion in the 3′ UTR affects mRNA stability. This
mechanism also has the potential to affect gene expression and phenotypes. While our study hints at a correlation with potential adaptive significance, future studies should explore whether such phenotypic diversification ultimately confers an adaptive advantage and whether TEs can drive adaptation.

Materials and Methods

More detailed information on the materials and methods used in this study is provided in SI Appendix, Supplementary Information Text. The C. rubella accessions were described in our previous study (14, 27, 32) or this study (SI Appendix, Table S4). Flowering time was assayed as days to flowering after planting. TEs in C. grandiflora (20) and C. rubella (20) were annotated using RepeatMasker v.4.0.6 (www.repeatmasker.org). The polytrophic TEs were identified using TEPID (44) based on the raw reads of 27 sequenced genomes of C. rubella accessions (18, 19) (SI Appendix, Table S1) with the MTE genome as the reference. A neighbor-joining tree of 28 C. rubella accessions was constructed using PHYLIp (version 3.696) (45). Ecological niche modeling was performed to measure niche similarity using ENMTools 1.3 (46). Pooled leaves of early-flowering 879 × MTE (844 × MTE) F2 plants were sequenced to map the causal loci using SHOREmap (47). For the complementation assay and the poly(A) signal mutation assay of FLOWERING LOCUS C (FLC) in Arabidopsis thaliana with FRT2 and fox-3 alleles in the Col-0 background. For the GUS-reporter assay, the promoter was amplified from MTE and cloned into the pB121 vector using the SbfI and XmaI sites and introduced into Arabidopsis thaliana pCAMBIA1300 vector using the KpnI and SalI sites and introduced into the plant genus Arabidopsis thaliana (Brassicaceae).

ACKNOWLEDGMENTS. We thank Detlef Weigel, John Bowman, Xiaofeng Cao, and Wolfgang Busch for valuable suggestions and discussions; Li Yang and Mao-Fu Li for assistance with the experiments; and especially the anonymous reviewers for their help in improving the manuscript. This study was funded by National Natural Science Foundation of China Grant 91731306 (Y.-L.G.); Strategic Priority Research Program of the Chinese Academy of Sciences Grant XDB27010305 (Y.-L.G.); the Innovative Academy of Seed Design, Chinese Academy of Sciences Grant XDB27010305 (Y.-L.G.); Chinese Academy of Sciences President’s International Fellowship for Visiting Scientists (S.B.); and Monash University Larkins Fellowship (to S.B.).

1. Scheffers BR et al. (2016) The broad footprint of climate change from genes to biomes to people. Science 354:719.
2. Estoup A et al. (2016) Is there a genetic paradox of biological invasion? Annu Rev Ecol Evol Syst 47:31–57.
3. Kolbe JI et al. (2004). Genetic variation during flowering by a Cuban lizard. Nature 431:177–181.
4. Frankham R (2005) Resolving the genetic paradox in invasive species. Heredity (Edinb) 94:385.
5. Stapley J, Santure AW, Dennis SR (2015) Transposable elements as agents of rapid adaptation may explain the genetic paradox of invasive species. Mol Ecol 24:2241–2252.
6. Schrader L et al. (2014) Transposable element islands facilitate adaptation to novel environments in an invasive species. Nat Commun 5:5495.
7. Casacuberta E, Gonzalez J (2013) The impact of transposable elements in environmental adaptation. Mol Ecol 22:1503–1517.
8. Chuong EB, Elde NC, Feschotte C (2017) Regulatory activities of transposable elements: From conflicts to benefits. Nat Rev Genet 18:71–86.
9. Barrón MG, Fiston-Laviger AS, Petrov DA, Gonzalez J (2014) Population genomics of transposable elements in Drosophila. Annu Rev Genet 48:561–581.
10. Li ZW, et al. (2018) Transposable elements contribute to the adaptation of Arabidopsis thaliana. Genome Biol Evol 10:2140–2150.
11. Studer A, Zhao Q, Ross-Ibarra J, Doebley J (2011) Identification of a functional transposon insertion in the maize domestication gene tb1. Nat Genet 43:1160–1163.
12. Martin A, et al. (2009) A transposon-induced epigenetic change leads to sex determination in melon. Nature 461:1135–1138.
13. van der Hof AE, et al. (2017) The industrial melanin mutation in British peppered moths is a transposable element. Nature 534:102–105.
14. 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 USA 106:5246–5253.
15. Foxe JP, et al. (2009) Recent speciation associated with the evolution of selfing in Capsella. Proc Natl Acad Sci USA 106:5241–5245.
16. Hurka H, Neuffer B (1997) Evolutionary processes in the genus Capsella (Brassicaceae). Plant Syst Evol 206:295–316.
17. Tang C, et al. (2007) The evolution of selfing in Arabidopsis thaliana. Science 317:1070–1072.
18. Ågren JA et al. (2014) Mating system shifts and transposable element evolution in the plant genus Capsella. BMC Genomics 15:602.
19. Wu G, et al. (2010) Long-term balancing selection contributes to adaptation in Arabidopsis and its relatives. Genome Biol 18:217.
20. Slotte T, et al. (2013) The Capsella rubella genome and the genomic consequences of rapid mating system evolution. Nat Genet 45:831–835.
21. Huq AE, et al. (2010) Long-term balancing selection contributes to adaptation in Arabidopsis and its relatives. Nature 461:1135–1138.
22. Wei L, Cao X (2016) The effect of transposable elements on phenotypic variation: Insights from plants to humans. Sci China Life Sci 59:24–37.
23. Hurka H, Friesen N, German DA, Franzeke A, Neuffer B (2012) ‘Missing link’ species Capsella orientalis and Capsella rutila elucidate evolution of model plant genus Capsella (Brassicaceae). Mol Ecol 21:1223–1238.
24. Weigel D (2012) Natural variation in Arabidopsis: From molecular genetics to ecological genomics. Plant Physiol 158:2–22.
25. Weigel D (2012) Natural variation in Arabidopsis: From molecular genetics to ecological genomics. Plant Physiol 158:2–22.