A New Family of Ty1-copia-Like Retrotransposons Originated in the Tomato Genome by a Recent Horizontal Transfer Event

Xudong Cheng,^{*,†} Dongfeng Zhang,[†] Zhukuan Cheng,[†] Beat Keller[§] and Hong-Qing Ling^{*,1}

^{*}The State Key Laboratory of Plant Cell and Chromosome Engineering, National Center for Plant Gene Research, Beijing 100101, China,

[†]Graduate University of the Chinese Academy of Sciences, Beijing 100049, China, [‡]The State Key Laboratory of Plant Genomics, Institute of Genetics and Developmental Biology, Chinese Academy of Sciences, Beijing 100101, China and [§]Institute of Plant Biology, University of Zürich, 8008 Zürich, Switzerland

Manuscript received November 26, 2008

Accepted for publication January 9, 2009

ABSTRACT

Rider is a novel and recently active Ty1-copia-like retrotransposon isolated from the T3238*fer* mutant of tomato. Structurally, it is delimited by a duplication of target sites and contains two long terminal direct repeats and an internal open reading frame, which encodes a Ty1-copia-type polyprotein with characteristic protein domains required for retrotransposition. The family of *Rider* elements has an intermediate copy number and is scattered in the chromosomes of tomato. *Rider* family members in the tomato genome share high sequence similarity, but different structural groups were identified (full-size elements, deletion derivatives, and solo LTRs). Southern blot analysis in Solanaceae species showed that *Rider* was a Lycopersicon-specific element. Sequence analysis revealed that among other plants, two *Arabidopsis* elements (named as *Rider-like 1* and *Rider-like 2*) are most similar to *Rider* in both the coding and noncoding regions. RT-PCR analysis indicates that *Rider* is constitutively expressed in tomato plants. The phylogeny-based parsimony analysis and the sequence substitution analyses of these data suggest that these *Rider-like* elements originated from a recent introgression of *Rider* into the tomato genome by horizontal transfer 1–6 million years ago. Considering its transcriptional activity and the recent insertion of the element into at least two genes, *Rider* is a recently active retrotransposon in the tomato genome.

LONG terminal repeat (LTR) retrotransposons are structurally similar to retroviruses that can move within a host genome from one chromosome position to another through an RNA intermediate by a “copy-and-paste” mechanism (KUMAR and BENNETZEN 1999). LTR retrotransposons are classified as a large group of mobile elements, which are distinct from DNA transposons that move as a DNA molecule. A typical complete LTR retrotransposon has an open reading frame (ORF) in the internal region flanked by two LTRs. Normally, the ORF is divided into two regions, *gag* and *pol*; the former encodes an RNA-binding protein *gag*, which forms a virus-like particle, and the latter encodes several proteins, such as protease, integrase, reverse transcriptase, and RNase H, which are required for cDNA synthesis and integration of cDNA into host chromosomes (BOEKE and CORCES 1989). On the basis of domain order within the *pol* protein and sequence similarity, LTR retrotransposons are further classified as either the Ty3-gypsy- (Metaviridae) or the Ty1-copia- (Pseudoviridae) type (KUMAR and BENNETZEN 1999). Both types of

retrotransposons are found in high numbers in the genomes of higher plants (from 5.5% of the *Arabidopsis thaliana* genome to >50% of the genomic content of *Zea mays*) (SANMIGUEL *et al.* 1996; ARABIDOPSIS GENOME INITIATIVE 2000). There is increasing evidence that LTR retrotransposons are a major driving force for genome evolution and that they contribute to genome organization as well as to the regulation of gene activity (FLAVELL *et al.* 1992). To date, a series of LTR retrotransposons have been isolated and characterized from a wide range of plant taxa from monocotyledonous to dicotyledonous plants. The Ty1-copia elements show a broad insertion pattern, heterogeneity, and sequence variability (KUMAR and BENNETZEN 1999). Some of these elements display an insertion bias toward coding regions (FLAVELL *et al.* 1992).

Most of the isolated Ty1-copia retrotransposons are nonfunctional due to the presence of stop codons, frameshift mutations, or deletions, and only a few Ty1-copia retrotransposons characterized from plants have been shown to be active, such as *Tnt1* and *Tto1* from tobacco, the *Tos* elements from rice, and *BARE-1* from barley (FLAVELL *et al.* 1992). The activity of these elements was detected only during certain stages of plant development or stress conditions, indicating that their transposition is regulated at the transcriptional level (FLAVELL *et al.* 1992). To better understand the

Sequence data from this article have been deposited with the EMBL/GenBank Data Libraries under accession no. EU195798.

¹Corresponding author: Institute of Genetics and Developmental Biology, Chinese Academy of Sciences, Datun Rd., Chaoyang District, Beijing 100101, China. E-mail: hqling@genetics.ac.cn

structure and diversity of LTR retrotransposons as well as their impact on genomes, it is necessary to isolate and characterize additional active elements. The discovery of those active LTR retrotransposons not only could lead to a better understanding of retrotransposon gene expression and transposition, but also could be useful as genetic markers (ELLIS *et al.* 1998).

Vertical or horizontal (cross-species) transfers are postulated as two alternative ways in which LTR retrotransposons evolve. Vertical transmission by descent from a common ancestor has been proposed as the main distribution mechanism of LTR retrotransposons (JORDAN and McDONALD 1998), while the data on horizontal transfer, the process by which DNA elements move between distinct organisms, have been less conclusive because the elements occurred too long ago or between such closely related species that we cannot completely eliminate the alternative vertical transmission hypothesis (JORDAN and McDONALD 1998; JORDAN *et al.* 1999). In prokaryotes, a significant proportion of their genetic diversity is based on the acquisition of sequences from distantly related organisms. The lateral gene transfer plays an integral role in the evolution of bacterial genomes and in the diversification and speciation of the enterics and other bacteria (OCHMAN *et al.* 2000). However, in eukaryotes the horizontal gene transfer is much less studied. The few best-documented examples involving the horizontal transfer of elements are P (SILVA and KIDWELL 2000), mariner (ROBERTSON *et al.* 1998), and a *cop* element (JORDAN *et al.* 1999) as well as other elements (SILVA *et al.* 2004) in animals. In plants, horizontal transfer of a transposable *Mu*-like element (MULE) between the *Setaria* and rice and a retrotransposon *RIRE1* within the genus *Oryza* are documented (DIAO *et al.* 2006; ROULIN *et al.* 2008).

In this article, we report the isolation and characterization of the *cop*-like element *Rider* from T3238*fer*, a spontaneous iron-inefficient mutant of tomato (BROWN and CHANEY 1971). Our results reveal that it is a new type and Lycopersicon-specific Ty1-*cop* retrotransposon. It might originate from the *Arabidopsis* genome by a horizontal transfer event and still contain transpositional activity.

MATERIALS AND METHODS

The cosmid library of T3238*fer* was constructed with a pWEB cosmid cloning kit following the manufacturer's instructions (EPICENTRE Biotechnologies): ~20 µg of genomic DNA were first sheared by passing it through a standard pipette tip. After repairing fragments to generate blunt ends, 30- to 40-kb DNA fragments were selected on a low-melting-point agarose gel by comparison with a supplied 36-kb DNA standard and purified according to the instructions of the manufacturer. Subsequently, the selected DNA fragments were ligated into the supplied blunt-ended Cloning-Ready pWEB cosmid vector, packaged using ultra-high efficiency Max-Plax Lambda Packaging Extracts (>10⁹ pfu/mg for phage λ), and plated on the

included EPI100-T1R phage T1-resistant *Escherichia coli* plating strain. The library screening was performed with an improved PCR-based library screening method (CHENG *et al.* 2004). Two pairs of gene-specific primers flanking the two ends of the insertion event of *FER* were designed and used for the library screening. They are left1 5'-GCTGCAATGTGTCGCCCTTT-3', left2 5'-GCTTTGCGATCCTTAGTTC-3', right1 5'-CCTCTCC TTTTGGCTCCGA-3', and right2 5'-GTTATGCATATTGGG CTTATTAAT-3'. The insertion fragment was sequenced by primer walking applied Biosystems ABI Prism 3700 DNA Analyzer and BigDye Terminator kit v3.1 (following the standard instructions of the manufacturer).

To identify the LTR and target-site duplication (TSD) sequences as well as the internal domains, the sequence was analyzed by comparison with itself and *Tnt1*-94 (X13777) with the Dotter program (SONNHAMMER and DURBIN 1995). The tRNA-Met sequence used for the identification of the primer binding site (PBS) locus was obtained from a tRNA database (LOWE and EDDY 1997) (<http://rna.wustl.edu/tRNAdb/>), and a search for conserved protein domains was carried out with RPS-Blast (MARCHLER-BAUER *et al.* 2003) (<http://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi>). The insertion times of LTR retrotransposons with both LTRs were determined in a manner similar to the method by MA *et al.* (2004). For each complete copy, the two LTRs were aligned using the GCG program, and the nucleotide substitution (transitions + transversions) rate between the two LTRs was estimated by the Kimura two-parameter method of MEGA software (indels and microsatellites were not taken into account in estimating these divergence rates) (KUMAR *et al.* 2004). LTR divergence rates were then converted into dates using the average substitution rate of 6.96×10^{-9} substitutions/synonymous site/year (MA *et al.* 2004). BLAST searches (ALTSCHUL *et al.* 1997) with a nucleotide sequence of the *Rider*-1 reverse transcriptase (RT) domain against public and local nucleotide databases were used for the identification of related elements with the threshold *E*-value of $E < 1 \times 10^{-10}$. The sequence of the conserved RT domain extracted from each identified element was defined as described by XIAO *et al.* (2008), and the RT neighbor-joining tree (Poisson correction) was constructed using MEGA 2.1 (KUMAR *et al.* 2004) based on ClustalX alignments. The GenBank accession numbers for the sequences used in the phylogenetic analysis are as follows: *Rider*, EU195798; *TY4*, M94164; *pCal*, AF007776; *Tdh5*, AJ439552; *Le.Copia*, AY678298; *TLC1.1*, AF279585; *Tnt1*, X13777; *Rider-like 1*, AL138663; *At.Copia2*, AF287471; *Ta1.1*, X13291; *Ta1.3*, X53973; *AtRE1*, AB021263; *AtRE2.1*, AB021266; *Os.Copia1*, NM_001072329; *Os.Copia2*, U72726; *Osr1*, AB046118; *Zm.Copia*, AF466202; *Stonor*, AF082134; *Ins2*, AF434192; *BARE*, Z17327; *Osser*, X69552; *CIRE1.1*, AM040263; *MIRE1*, AY196987; *Tgm*, U96748; *Ta.Copia*, DQ890165; *Tar1*, AB008772; *Ib.Copia*, AY830138; *Pb.Copia*, AJ416708; *CIRE1.1*, AM040263.

To perform the fluorescence *in situ* hybridization (FISH) and fiber-FISH, immature tomato flower buds of ~3.0 mm were harvested and fixed in Carnoy's solution (ethanol:glacial acetic = 3:1). Microsporocytes at meiosis were squashed in acetocarmine solution as described by WU (1967). Slides were frozen in liquid nitrogen. After removing coverslips, they were dehydrated in an ethanol series (70, 90, and 100%) prior to use in FISH. The FISH procedure used for chromosomes was performed according to JIANG *et al.* (1995). A *Hind*III restriction fragment that contained the LTR and gag-protease region of *Rider* was used as *Rider*-specific probe and labeled by nick translation with digoxigenin-16-dUTP (Roche). After hybridization and washing, chromosomes and FISH signal images were captured under the Olympus BX61 fluorescence microscope conjunct with a microCCD camera.

Genomic DNA extraction was performed as described by Li *et al.* (2004). For Southern blot analysis, ~10 µg genomic DNA was separately digested with *Eco*RI or *Hind*III. Gel separation, blotting of DNA fragments onto Hybond-N⁺ membrane (Amersham), and hybridization were performed according to the protocol described by Ling *et al.* (1996). The probes were synthesized by PCR amplification and labeled with ³²P by the Prime-a-Gene System (Promega).

Total RNA was isolated from leaves and roots of 3-week-old T3238 and T3238*fer* plants and flowers as described by Li *et al.* (2004) and treated with RNase-free DNase I (Promega) to eliminate genomic DNA, and then 2 µg/sample was used for reverse transcription with the M-MLV RT kit (Promega) according to the manufacturer's instructions. The reactions were diluted sixfold for RT-PCR analysis according to the protocol described by Li *et al.* (2004). The internal fragment-specific primers used in the RT-PCR experiments are internal-forward and internal-reverse as described above (real-time PCR analysis). The other primers used in this experiment are p-l 5'-GGGCCGACGGTATACAATG-3', p-r 5'-ACCGAGAGGC TCTGATACCA-3', and *fer* 5'-CAAAGGCACGAGGACTGACC-3'. The PCR reaction profile included 26 cycles of 30 sec at 94°, 50 sec at 55°, and 50 sec at 72°, preceded by an initial denaturation (3 min at 94°) and followed by a final extension step (5 min at 72°). Reaction products were separated on agarose gel electrophoresis.

For monitoring the real-time PCR reactions, we used the Mastercycler ep realplex system (Eppendorf) and DyNamo SYBR Green quantitative PCR (qPCR) kit for fluorescence labeling (New England Biolabs) with optimized PCR protocols. The PCR reaction system (20 µl) contained 10 µl DyNamo SYBR Green qPCR mix, 0.5 µM forward and reverse primer, and 2 µl DNA template (0–100 ng). The standard amplification profile consisted of 2 min at 95° for initial denaturation and 40 cycles of amplification of 10 sec at 95° for denaturation, 30 sec at 60° for annealing/extension, and 20 sec at 72° with a single fluorescence measurement. Melt curve analysis was performed subsequent to the PCR run to examine the specificity of amplification. After adjusting baseline cycles and calculating threshold values, the number of cycles halfway through the exponential phase (Ct value) was obtained. The relative standard curve methods were used for quantification of copy number with a 10-fold dilution series of 4.8×10^{-9} copies/µl of *Rider* as plasmid DNA. The genome sample was genomic DNA of T3238 (5.0 µg/µl) and water as negative control. When converting the result of relative copy number into absolute copy number, we used a tomato genome size of 960 Mb (Van der Hoeven *et al.* 2002). The quantification of each sample was repeated three times. To amplify LTR and internal fragments, the following PCR primer pairs were used: LTR-f, 5'-TGAATCGGACCCGCTACAA-3'; LTR-r, 5'-TGACCG GGCAGCGAGA-3'; internal-f, 5'-TGTCCTAGGAGAGAGTG GTTC-3'; and internal-r, 5'-CATTAGAGAGCATGCACCTTG-3'.

RESULTS

Isolation and sequence analysis of the Ty1-copia-like retrotransposon *Rider*: T3238*fer*, a spontaneous iron-inefficient mutant of tomato, was found in the 1960s (Brown and Chaney 1971). Our previous work has demonstrated that the mutation of T3238*fer* was caused by an insertion of a large DNA fragment in the first exon of *FER*, blocking the expression of this gene that functions in controlling iron deficiency responses and iron uptake (Ling *et al.* 2002). T3238*fer* is a lethal mutant

under normal culture conditions because of its inefficiency in acquiring iron from the soil. Using a PCR-based library screening method (Cheng *et al.* 2004) and the specific PCR primers of *FER*, a cosmid clone covering the whole insertion was isolated from a genomic cosmid library of T3238*fer* and sequenced by primer walking. Finally, a 6-kb sequence, which contained the complete insertion and the exon of *FER*, was obtained.

Using the dotplot program (Sonnhammer and Durbin 1995), the insertion fragment in *FER* was identified to be 4876 bp long and flanked by two 5-bp short directed repeats (5'-CTTTT-3'). Adjacent to the short directed repeats there are two 397-bp LTRs that terminate by the consensus sequence 5'-TG...CA-3', and a 3921-bp internal putative coding region (Figure 1A). The short directed repeat is a typical TSD, which is a characteristic feature of a transposon insertion event. On the basis of the identification of LTR sequences, the inserted element in *FER* possibly belongs to the class of LTR retrotransposons.

In the internal region between the LTRs, two conserved sites, the primer binding site (PBS) and the polypurine tract site (PPT) (Figure 1A), were identified 3 bp downstream from the left LTR and closely upstream from the right LTR, respectively. The PBS is part of a conserved 5'-TGGTATCAGAGCC-3' sequence, which serves to complement the 3' sequence of tRNA (Met) and to activate the transcription of mRNA (Akama and Tanifuji 1989). The polypurine tract site with the conserved AGGGGAG motif (Friant *et al.* 1996) is speculated to prime the synthesis of second-strand DNA. Between the two sites, a large single ORF was identified. It encodes a polyprotein composed of 1307 amino acids. Searching for conserved domains using RPS-Blast allowed us to identify gag, zinc-finger, protease (pro), integrase (int), and RT domains within the polyprotein together with the domain of ribonuclease H (RH). In comparison with all published LTR elements, we found that the general organization of these domains was very similar to Ty1-copia retrotransposons both in order and in length. A homology matrix comparison of the deduced amino acid sequences between this element and tobacco *Tnt1* is shown in Figure 1B. The identity value at the protein level between the two ORFs was 45% and the two polyproteins showed a common order of gag, pro, int, RT, and RH domains. Whereas different domains showed different levels of sequence similarity, the protease, integrase, and reverse transcriptase-RNase H domains were most conserved with identities of 43% in 98 amino acids, 50% in 158 amino acids, and 57% in 599 amino acids, respectively. The gag polyprotein is less conserved with 40% identity in 246 residues. This high degree of conservation in amino acid sequence and domain order indicates that the tomato element contains all characteristic protein domains required for retrotransposition, suggesting that this element belongs to a new Ty1-copia family of LTR retrotransposons. Most

FIGURE 1.—Schematic of *Rider* and the sequence comparison. (A) Structure of the full-length *Rider* element. LTRs are depicted as shaded arrows, the ORF is represented by a solid box, and positions of stop codons in three reading frames are shown as vertical lines below the structure of *Rider*. Nucleotide sequences of TSD, LTR ends, PBS, and PPT are shown above the structure. (B) Homology matrix comparison between the *Rider* element and the *Tnt1* internal coding region. The computer program of Dotter (FRIANT *et al.* 1996) was used to generate diagonal lines indicating segments of 20 residues that showed similarity above a threshold level. The nucleotide sequence of *Rider* was first translated into protein sequences within the three open reading frames and compared with the polypeptide of *Tnt1*. Nucleotide and residue numbers are indicated. The approximate regions corresponding to the gag protein and the endonuclease (int), RT, and RH domains encoded by the *pol* gene are indicated below the dotted lines.

recently, XIAO *et al.* (2008) described a gene duplication underlying the morphological variation of tomato fruit mediated by the retrotransposon *Rider*, whose sequence is completely identical to our isolated element. Thus, here we call our isolated element *Rider*, too.

***Rider* elements in the tomato genome:** To identify sequences homologous to *Rider*, we searched all the published tomato BAC or PAC sequences using the full-length sequence of *Rider* as the query. A set of 32 matches (>200 bp and identity value >90%) were identified from 25 tomato bacterial artificial chromosome (BAC) or P1 artificial chromosome (PAC) clones distributed on chromosomes 1, 4, 5, 8, and 10. These matches were highly similar in sequence and differed from any retrotransposable elements previously isolated in tomato. Among these matches, 6 are full-size homologs with the same overall structure as *Rider*; 4 are putative nonautonomous elements that lost partially or completely internal coding regions, and 22 are solo LTRs or deletion derivatives. Such divergent structural groups indicate that frequent rearrangement, deletions, or template switching events occurred in the *Rider* family of retrotransposons.

We analyzed the six full-size *Riders* individually (Table 1), revealing that they all have completely conserved PBS or PPT sites and a single internal ORF (*Rider-3* and *Rider-6* have one or two stop codons in the gag–pro regions). Distinct TSDs could be identified, too, except *Rider-5*, revealing that the TSD sequence of each element was different, *e.g.*, 5'-CTTTT-3' in *Rider-1* and 5'-CTAGC-3' in *Rider-2* (Table 1). This suggests that integration sites of *Rider* in tomato chromosomes might be randomly distributed. As the two LTRs are created by a duplication of a single template during replication of the element, followed by the accumulation of random mutations, the degree of their divergence is proportional to the time passed since the insertion of the element. Therefore, the nucleotide divergence between the two LTRs of each full-size element was compared and used to estimate the insertion time. As shown in Table 1, five elements contained few mutations (>98% identity) between the two LTRs. *Rider-6* had six nucleotide changes in the 397-bp LTR sequence and was determined to be the most ancestral insertion (1.07 MYA). *Rider-1*, *Rider-2*, and *Rider-7* were the three youngest elements. All had

TABLE 1
Features of *Rider* members in the tomato genome

Member	Locus: BAC or YAC	TSD 5'–3'	LTR identity (%)	LTR K ($\pm$ SE)	Insertion time (MYA)	Stop codon in ORF
<i>Rider-1</i>	56B23/chr 6	CTTTT	100	0.000 (0.000)	0.00	None
<i>Rider-2</i>	AP009278/chr 8	CTAGC	100	0.000 (0.000)	0.00	None
<i>Rider-3</i>	AC194694/chr 6	CATTT	99.3	0.007 (0.005)	0.50	2
<i>Rider-4</i>	AC171727/chr 1	GACAA	99.2	0.008 (0.004)	0.57	None
<i>Rider-5</i>	AC204082/chr 5	—	98.8	0.013 (0.006)	0.93	None
<i>Rider-6</i>	AC204082/chr 5	TATAT	98.2	0.015 (0.006)	1.07	1
<i>Rider-7</i>	EF094940/chr 10	GACCT	100	0.000 (0.000)	0.00	None
<i>Rider-l1</i>	AL138663/chr 3	—	93.0	0.068 (0.017)	4.89	Many
<i>Rider-l2</i>	AC007188/chr 2	TCTCC	91.0	0.078 (0.019)	5.60	Many

LTR identity percentage was calculated with GAP (GCG). K is the estimate of the number of substitutions per nucleotide site and its standard error is shown; they were calculated with MEGA 3.0 (KUMAR *et al.* 2004) on the basis of the K2P model. Insertion time of divergence was calculated on the basis of the average substitution rate (r) of 6.96×10^{-9} substitutions/synonymous site/year for tomato (TANKSLEY *et al.* 1992). *Rider-7* was identified and described by XIAO *et al.* (2008), *Rider-like 1* (*Rider-l1*) and *Rider-like 2* (*Rider-l2*) are identified in the genome of Arabidopsis. chr, chromosome.

completely identical LTRs and their insertion must have occurred very recently. This is consistent with the occurrence of the T3238*fer* mutation caused by *Rider-1* in the 1960s.

When using the *Rider* element as the query sequence of BLASTN, another interesting match, GTOM5, was obtained from the nucleotide database of NCBI. The GTOM5 mRNA (X67143) is an aberrant transcript of the phytoene synthase gene (*Psyl*) that was isolated from the tomato mutant *yellow flesh*. This sequence contains the first 326-bp nucleotides of the *Psyl* gene followed by 181 bp of an inserted sequence. Southern hybridization analysis had shown that this inserted sequence was highly repetitive in the tomato genome (FRAY and GRIERSON 1993). Blast analysis revealed that the repeated sequence was similar to part of the *Rider* LTR, sharing 96% identity at the nucleotide level. FRAY and GRIERSON (1993) reported that the insertion of the repeated segment was the reason for the transcriptional lesion of the phytoene synthase gene resulting in the yellow-fruited mutant *yellow flesh* (PRICE and DRINKARD 1908). Therefore, this mutant is most likely another case of a *Rider* mutational insertion event in the tomato genome.

Four *Rider* nonautonomous elements were also analyzed. The ones identified from the sequences of AJ439079, AP009276, and AC18878, respectively, showed degraded TSD sites, and two of them from AP009276 and AC188781 could still be identified to contain the intact PBS and PPT sites, whereas the TSD and PBS/PPT sites of the fourth, which was identified from the 5' upstream region of tomato gene *LEACO1* (X58273) and described as a typical repeated sequence with multiple copies in the tomato genome (BLUME *et al.* 1997), were completely degenerated.

Transcriptional activity analysis: The tomato EST database (<http://sgn.cornell.edu>) and NCBI EST database were searched for evidence of the transcriptional

activity of *Rider*. In tomato, seven EST hits with >95% identity were obtained. The lengths of these EST hits varied from 433 to 808 bp. Three of them are located in the RT–RH domain and four are in the LTR. Then we experimentally analyzed the transcriptional activity of the *Rider-1* that inserted in the *FER* gene. As mentioned above, the aberrant transcript GTOM5 was caused by a *Rider* insertion and the expression of this RNA was under the control of the *Psyl* promoter (FRAY and GRIERSON 1993). To test the *Rider* transcription driven by the *FER* promoter, two *FER* gene-specific primers (left-1 and right-1) designed from the 5' and 3' regions of the first exon of *FER* were used in combination with the *Rider*-specific primer LTR-r located in the reverse direction of the left LTR or other internal and LTR primers (Figure 2A), respectively. Using left-1/LTR-r primers for RT–PCR, a very weak band was detected in the root sample of T3238*fer* with 40 PCR cycles, while no products were identified with the other primer pairs (Figure 2B). This is consistent with the root-specific transcriptional activity of *FER* (LING *et al.* 2002) and indicates that the band might be due to *FER* promoter activity. However, this transcription was at a very low level and could not extend into the internal region of *Rider* due to a stop signal in the LTR regions.

The RT–PCR amplification with the primers of LTR-l/LTR-r and internal-l/internal-r, derived from LTR and gag–protease sequences (Figure 2A), respectively, was performed with RNAs from leaves, roots, and flowers of T3238 and the T3238*fer* mutant. As shown in Figure 2, positive PCR products were amplified from roots, leaves, and flowers of T3238*fer*. Sequencing analysis verified that the sequences of the PCR products were 98–100% similar to *Rider-1*, which indicated that they were from the transcription of *Rider-1* or its homologs. RNA samples from plants grown under stress conditions such as drought or tissue culture were also analyzed, but no differences in expression were found (data not shown).

FIGURE 2.—Transcriptional activity of *Rider* in T3238 and T3238*fer* analyzed by RT-PCR. (A) Schematic of *fer* locus with the *Rider* element in T3238*fer*. The primer pairs used for *Rider* transcription activity analysis are indicated by short arrows. (B) RT-PCR analysis of *Rider* transcription in roots, leaves, and flowers of T3238 and T3238*fer*. A plasmid containing *Rider* is used as the positive control. The expression of the elongation factor *Leaf-1A* was monitored as a control for RT-PCR.

Copy-number estimation and chromosome distribution: With the primers specific to LTR and protease domain sequences, a quantitative real-time PCR amplification was performed to estimate the copy number of

Rider elements in the tomato genome. The relative standard curve was made by a 10-fold dilution series of plasmid DNA containing *Rider*. As shown in Figure 3A, the slopes of the two five-point standard curves, $C_t = f(-\log[\text{DNA}])$, were -1.954 and -0.652 , respectively, with a correlation index >0.95 for both curves. From the standard curves, the number of LTRs and internal regions of *Rider* in the T3238 genome were determined and are shown in Figure 3A (right): a total of 195 LTR copies and 66 protease copies were estimated in the T3238 genome. The ratio calculated from copy numbers of LTR *vs.* protease regions was >2 . Considering that each *Rider* element has one copy of the internal region and two LTRs, the ratio suggests that many solo *Rider* LTRs exist in the tomato genome.

To study the global distribution of *Rider* members on tomato chromosomes, a fiber-FISH analysis was carried out with the fragment containing both the LTR and gag-protease regions as probe. The *in situ* hybridization showed that the *Rider* members were abundant and scattered in tomato chromosomes (Figure 3B). This localization differs from most plant Ty1-*copia* group elements, which are euchromatic (KUMAR *et al.* 1997).

FIGURE 3.—Estimation of *Rider* copy number and distribution in tomato genome. (A) Copy-number estimation with qPCR. (Left) qPCR standard curves (plotting template amount *vs.* C_t value) from a serial dilution of standard DNA and amplification with internal/LTR-specific primer (see MATERIALS AND METHODS for details). The correlation coefficient values indicate high linearity (r^2) and PCR efficiency (E). (Right) The estimated copy number of internal and LTR regions in the T3238 genome are given. Columns present copy numbers derived from relative ΔC_T values (refer to the manual of Mastercycler ep realplex system). Error bars reflect the difference between replicate PCRs using the same DNA sample. (B) FISH mapping of *Rider* in tomato chromosomes with a probe containing an LTR and the polypeptide coding region. (Left) A gray-scale image of tomato chromosomes. (Right) A fluorescence image of the same field. The yellow points indicate the locations of *Rider* elements.

FIGURE 4.—Southern blot analysis of several species of Solanaceae with specific probes derived from internal (middle) and LTR (right) regions of *Rider*. The total genomic DNAs were digested with *Eco*RI. The phylogenetic tree on the left is from NÚEZ *et al.* (2004). The direction of DNA migration on the gel is indicated by the dashed arrow. *C. arabica*, *N. tabacum*, and *S. tuberosum* are coffee, potato, and to-

bacco, respectively, and *L* means *Lycopersicon*. The thick arrow indicates the time point when the *Rider* element was integrated into the ancient genome of tomato by a horizontal transfer.

***Rider* elements in other species of the Solanaceae family:** To check the distribution of the *Rider* retrotransposon family in Solanaceae, six *Lycopersicon* species (*L. hirsutum*, *L. chmielewskii*, *L. peruvianum*, *L. chilense*, *L. pimpinellifolium*, and *L. esculentum*) and three distant relatives (coffee, potato, and tobacco) were analyzed by Southern blot hybridization together with T3238fer. As shown in Figure 4, after hybridization, signals were observed in all *Lycopersicon* species even when washed by high washing stringency (2× SSC and 0.1% SDS), whereas no hybridization signals were detected in coffee, potato, and tobacco even at a low washing stringency (2× SSC and 0.5% SDS) with the probes derived from the LTR region and the internal gag domain of *Rider*. These results suggest that the *Rider* retrotransposon family is specific in *Lycopersicon* species. Additionally, the intensity of hybridization signals varied among the species. With both probes, *L. chilense* displayed the weakest hybridization signals among the species investigated, whereas *Rider* elements are more abundant in *L. esculentum* (Figure 4).

To further confirm the presence of *Rider*-retrotransposable elements in other species, we carried out a BLASTN search in the nr/nt database of NCBI with the internal coding region of *Rider* as the query sequence. A total of 156 matches, distributed in 14 genomes and sharing 40% of the common region with the query sequence, were obtained (Figure 5). All the 156 nucleotide sequences were extracted for alignment with the full-size *Rider* sequence individually, and the regions of nucleotide identities aligned to *Rider* were drawn. Figure 5B (red indicates that the identity score is high) illustrates the regions of nucleotide identities among the representative elements from the 14 species aligned to *Rider*. All of these elements only partially matched *Rider* in internal regions such as int, RT, or RH domains, except two putative Ty1-copia-like elements from Arabidopsis BAC clone AL138663 and AC007188. The two putative Arabidopsis elements matched *Rider* in both coding and noncoding regions with 75% identity at the

nucleotide level (Figure 6) and were named *Rider-like 1* and *Rider-like 2*, respectively. They are two complete retrotransposable elements with two 248-bp long LTRs terminated by the consensus sequence 5'-TG...CA-3', which share high sequence identity to the 3' part of *Rider* LTRs (Figure 6B). The PBS/PPT sites of the two *Rider-like*'s are identical to *Rider* also. Moreover, *Rider-like 2* has a pair of distinct TSD sequences (TCTCC) whereas the TSD sequence of *Rider-like 1* is degenerated. LTR-nucleotide divergence analysis of *Rider-like 1* and *Rider-like 2* showed that their LTRs share 93 and 91% identity, respectively. It allows us to estimate that the insertion time of *Rider-like 1* and *Rider-like 2* is ~4.89 and 5.60 MYA, respectively (Table 1). This result suggests that the "birth" of *Rider-like 1* and *Rider-like 2* in the Arabidopsis genome occurred after the radiation of tomato and potato genomes 12 MYA (TANKSLEY *et al.* 1992). While in the Arabidopsis genome, there are only two copies of *Rider-like*, no transcriptional products were detected by both EST database searching and RT-PCR amplification with *Rider-like*-specific primers (data not shown), suggesting that *Rider-like 1* and *Rider-like 2* are transcriptionally inactive.

Phylogenetic analysis: To gain insight into the evolutionary relationship of *Rider* and other copia-type retrotransposons, the sequences of their conserved RT domains were extracted from 24 previously identified copia-type elements as well as from *Rider* and its homologs (from *Oryza sativa* and *Triticum aestivum*, as shown in Figure 5B) for the phylogenetic tree reconstruction. As shown in Figure 7, the unrooted neighbor-joining tree revealed that the various RT sequences clustered into several distinct major evolutionary lineages supported by high bootstrap value and topographic structure. The long branches and loose lineage structure indicate a considerable diversity and similar time of origin of each group, consistent with the previous classification of plant copia families (WICKER and KELLER 2007). In the tree, *Rider* was clustered into the group with *Rider-like 1* of Arabidopsis, *Os.Copia1*, and *Ta.Copia*, but not with the

FIGURE 5.—Analysis of *Rider* sequences. (A) Phylogeny of plants by HEDGES (2002). The Solanaceae–Brassicaceae divergence time is presented as a range (100–120 MY) on the basis of WICKER and KELLER (2007). A time estimation of 12 MY for the divergence of *L. esculentum* (tomato) and *S. tuberosum* (potato) is also shown, referring to TANKSLEY *et al.* (1992). Red line branches indicate that the species contain some homologous element(s) with *Rider*. The thickness of red lines implies the degree of homology. (B) Schematic of the regions with nucleotide identities aligned to *Rider*. The overall structure of the *Rider* element is indicated at the top as a black box with gray arrows to depict LTRs and vertical white lines to divide the internal domain regions. All the homologous sequences of *Rider* (sharing more than 40% identical region, <1E-10) in the nr/nt database of NCBI are identified by BLASTN search

and the most similar one within a plant is drawn to scale; gray boxes indicate regions that are conserved with *Rider*. Note that the Arabidopsis homolog, *Rider-like*, is well conserved with *Rider* not only in the internal region but also in LTR. The GenBank accession numbers of the involved sequences are AM473609, AL138663, AC187141, AC190836, AC166742, AP004495, AC174796, AF47373, AC127428, DQ890165, AP008218, AC186755, AC196852, and AC198352. The red triangle between *A. thaliana* and *L. esculentum* implies the transferring direction of *Rider* element.

published elements from Solanaceae (shaded dark green in Figure 5). This is consistent with previous BLASTN searching as shown in Figure 5.

DISCUSSION

Sequence analysis revealed that the *Rider* element has all the structural features characteristic of Ty1-*copia*-type retrotransposons. Its LTR is completely different from any other known LTR element, whereas the internal regions share common sequences at a low identity level with the putative Ty1-*copia*-like elements identified from the plant genomes with the exception of *Rider-like 1* and *Rider-like 2* of Arabidopsis. Furthermore, *Rider* showed a constitutive transcriptional activity in all tissues investigated (Figure 2), and two recent mutational insertions were observed in *FER* of the tomato mutant T3238*fer* (BROWN and CHANEY 1971) and *Psyl* of the mutant *yellow flesh* (FRAY and GRIERSON 1993). More recently, XIAO *et al.* (2008) found a *Rider* (*Rider-7*)-mediated gene duplication in the tomato genome that resulted in morphological variation of the fruit. These results strongly support *Rider* as a novel type of Ty1-*copia* retrotransposable element that was recently actively transposing in the tomato genome.

Vertical inheritance and horizontal transfer are considered as two main mechanisms for the evolution of a

transposable element. Southern hybridization and homology search analysis revealed that *Rider* is specific to Lycopersicon species (Figure 4) and is virtually identical to the Ty1-*copia*-like elements (*Rider-like 1* and *Rider-like 2*) of Arabidopsis (Figures 5 and 6). The noncoding region, including the LTR and PBS/PPT sites, which are expected to evolve more rapidly than the coding region of a retrotransposon, even showed high levels of identity between *Rider* of tomato and *Rider-like* of Arabidopsis.

The fact that *Rider* elements are absent in the genomes of related species of Lycopersicon such as potato, tobacco, and coffee (Figure 4) might be interpreted as a loss of *Rider* in these species in the framework of the vertical inheritance. However, this inference suffers from two serious problems. The first problem is that one has to assume many independent losses of the *Rider-like* elements after Solanaceae diverged from the most recent ancestor shared with Arabidopsis. For example, in the phylogeny shown in Figure 4, three independent losses have to be invoked in the lineages toward *Coffea arabica*, *Nicotiana tabacum*, and *Solanum tuberosum* to ensure that the *Rider-like* elements exist in every ancestor of these species and eventually passed to the tomato genus. The horizontal transfer hypothesis seems much more likely because it assumes only one introgression event into the ancestral tomato genome. This phylogeny-based parsimony analysis strongly supports the

FIGURE 6.—Comparison of *Rider* and *Rider-like* sequences at the nucleotide level. (A) Homology matrix comparison between the full-size *Rider* element and *Rider-like* elements. The computer program of Dotter (SONNHAMMER and DURBIN 1995) was used to generate diagonal lines indicating segments of 20 nucleotides that showed similarity above a threshold level. The LTR is indicated by solid arrows. (B) Nucleotide alignment of the LTR-PBS region of *Rider* and *Rider-like*. Positions of nucleotide identity are shaded.

horizontal transfer hypothesis. The second problem of the vertical transfer hypothesis is that, if this were the case, the age of the *Rider* family would be at least as old as the common ancestor of Solanaceae and Arabidopsis, which is known to have diverged 100–120 MYA (LONG *et al.* 1996). According to an evolutionary rate of $5.9\text{--}6.96 \times 10^{-9}$ substitutions at synonymous site per year in plants (WHITE and DOEBLEY 1999; MA *et al.* 2004), the silent sequences (nonprotein sequences) in the *Rider* members in the two branches should keep only 31–41% of identity, which is significantly lower than what we observed—~75%. On the basis of this analysis, the high identity with the *Rider* homologs in the distantly related species Arabidopsis (Figure 5) also supports that *Rider* originates from a horizontal gene transfer.

Rider-like elements in Arabidopsis are ~5.6 MY old, obviously older than the most ancient *Rider* element detected in tomato (with an age of 1.07 MY; Table 1), but younger than the speciation of tomato (~12 MYA) (TANKSLEY *et al.* 1992). Considering that *Rider-like* in Arabidopsis is inactive and largely unalignable at the 5' half of the LTRs to the tomato *Rider*, we infer that *Rider-like* of Arabidopsis and *Rider* of tomato are all from an ancestral *Rider* element of some genomes in Arabidopsis or related species by horizontal transfer that occurred 1–5.6 MYA. The transfers should be mediated via a vector of some kind, such as viruses, bacteria, fungi, and sap-sucking insects (SILVA *et al.* 2004; DIAO *et al.* 2006; ROULIN *et al.* 2008), and its transpositional activity was retained in the *Lycopersicon* genome, but lost in Arabidopsis.

FISH mapping and the determination of copy number demonstrated that the tomato genome contains an intermediate number of *Rider* homologs (~195 LTR copies and 66 internal copies) scattered in the genome of tomato. This distribution is consistent with the fact that the TSD sites among different element members are nonconserved. This feature differs from other plant Ty1-copia elements (FLAVELL *et al.* 1992; KUMAR *et al.* 1997). A more detailed knowledge on the distribution and the integration sites of *Rider* elements will be gained from the International Tomato Sequencing Project (http://www.sgn.cornell.edu/about/tomato_sequencing.pl).

Southern hybridization analysis also revealed many insertion polymorphisms of *Rider* among cultured tomato and its closely related wild species. This reveals a potential use of *Rider* for taxonomic studies. Several marker systems based on retrotransposons have been developed for plant genome and biodiversity analysis (PURUGGANAN and WESSLER 1995), and techniques allowing the PCR analysis of retrotransposon insertion polymorphisms have recently extended the application of retrotransposons as genetic markers (WAUGH *et al.* 1997; ELLIS *et al.* 1998). Therefore, studying the distribution of *Rider* in more tomato species could be useful for evaluation and clarification of the phylogenetic relationships in *Lycopersicon* genus.

The transposition of retrotransposons is believed to be regulated mainly at the transcriptional level. Most retrotransposons reported are silent or active only during certain stages of plant development (KUMAR *et al.* 1997) or under specific stress conditions (KUMAR *et al.* 1997;

FIGURE 7.—Phylogenetic analysis based on the alignment of RT sequences. The neighbor-joining tree was constructed with MEGA3.0 (KUMAR *et al.* 2004) on the basis of the alignment of nucleotide sequences extracted from the RT regions of *Rider* and 27 identified *copia*-type retrotransposons. The bootstrap values, shown at the nodes, are percentages for 1000 replications. The positions of *Rider* and *Rider-like* in the tree are red and the other four published elements from Solanaceae clade are blue. GenBank accession numbers used for RT sequence extraction are as follows: *Rider*, EU195798; *TY4*, M94164; *pCal*, AF007776; *Tdh5*, AJ439552; *Le.Copia*, AY678298; *TLC1.1*, AF279585; *Tnt1*, X13777; *Rider-like*, AL138663; *At.Copia2*, AF287471; *Ta1.1*, X13291; *Ta1.3*, X53973; *AtRE1*, AB021263; *AtRE2.1*, AB021266; *Os.Copia1*, NM_001072329; *Os.Copia2*, U72726; *Osr1*, AB046118; *Zm.Copia*, AF466202; *Stonor*, AF082134; *Ins2*, AF434192; *BARE*, Z17327; *Oser*, X69552; *CIRE1.1*, AM040263; *MIRE1*, AY196987; *Tgm*, U96748; *Ta.Copia*, DQ890165; *Tar1*, AB008772; *Ib.Copia*, AY830138; *Pb.Copia*, AJ416708; *CIRE1.1*, AM040263.

BEGUIRISTAIN *et al.* 2001; RICO-CABANAS and MARTINEZ-IZQUIERDO 2007). Such limited activity might minimize the deleterious effect of element amplification in the genome (KUMAR and BENNETZEN 1999). *Rider* showed a constitutive transcription in all tested tissues of tomato under different growth conditions and an intermediate number (~100 copies) in the tomato genome, which does not correlate with its transcription activity. Four steps (transcription, translation, reverse transcription, and integration of cDNA element) are included in a replication cycle of LTR retrotransposon. Regulation at any of the four steps can limit the transposition rate. The constitutive expression of *Rider* indicates that no amplification suppression exists at the transcriptional level. The regulation of the *Rider* transposition rate might occur after the transcription. Some constitutively expressed LTR retrotransposons have been observed in other plants (NEUMANN *et al.* 2003).

In conclusion, *Rider* is a novel, young, and recent active Ty1-*copia*-like retrotransposon in the tomato genome and

may originate from horizontal gene transfer. This finding is in agreement with earlier conclusions that the movement and amplification of retrotransposons are major contributors to genome evolution and genetic diversity in plant (JIN and BENNETZEN 1994; BERGTHORSSON *et al.* 2003; LONG *et al.* 2003; WANG *et al.* 2006). The possible hijacking of nuclear genes by retrotransposons could transfer them between organisms, leading to the origination of a new gene or a new gene recombination, representing a drastic innovation in a genome (JIN and BENNETZEN 1994; WANG *et al.* 2006). Further characterization of the *Rider* family will give new insight into the role of these retrotransposons in genome evolution and genetic diversity of tomato species.

The authors are grateful to Barbara Hohn (The Friedrich Miescher Institute for Biomedical Research, Basel, Switzerland) for critical reading of the manuscript. This work was supported by the Chinese Academy of Sciences (grant nos. KSCX2-YW-N-001 and KSCX2-YW-N-056), the Ministry of Science and Technology of China (grant nos. 2005cb20904 and 2006AA10A105), and the National Natural Science Foundation of China (grant no. 30530460).

LITERATURE CITED

- AKAMA, K., and S. TANIFUJI, 1989 Nucleotide sequence of a methionine initiator tRNA gene of *Arabidopsis thaliana*. *Plant Mol. Biol.* **13**: 599–600.
- ALTSCHUL, S. F., T. L. MADDEN, A. A. SCHAFER, J. ZHANG, Z. ZHANG *et al.*, 1997 Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* **25**: 3389–3402.
- ARABIDOPSIS GENOME INITIATIVE, 2000 Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. *Nature* **408**: 796–815.
- BEGUIRISTAIN, T., M. A. GRANDBASTIEN, P. PUIGDOMENECH and J. M. CASACUBERTA, 2001 Three Tnt1 subfamilies show different stress-associated patterns of expression in tobacco: consequences for retrotransposon control and evolution in plants. *Plant Physiol.* **127**: 212–221.
- BERGTHORSSON, U., K. L. ADAMS, B. THOMASON and J. D. PALMER, 2003 Widespread horizontal transfer of mitochondrial genes in flowering plants. *Nature* **424**: 197–201.
- BLUME, B., C. S. BARRY, A. J. HAMILTON, M. BOUZAYEN and D. GRIERSON, 1997 Identification of transposon-like elements in non-coding regions of tomato ACC oxidase genes. *Mol. Gen. Genet.* **254**: 297–303.
- BOEKE, J. D., and V. G. CORCES, 1989 Transcription and reverse transcription of retrotransposons. *Annu. Rev. Microbiol.* **43**: 403–434.
- BROWN, J. C., and R. L. CHANEY, 1971 Effect of iron on the transport of citrate into the xylem of soybeans and tomatoes. *Plant Physiol.* **47**: 836–840.
- CHENG, X. D., L. L. DONG and H. Q. LING, 2004 Improvement on PCR-based cDNA and genomic library screening. *Chinese H. Tech. Lett.* **14**: 32–37.
- DIAO, X., M. FREELING and D. LISCH, 2006 Horizontal transfer of a plant transposon. *PLoS Biol.* **4**: e5.
- ELLIS, T. H., S. J. POYSER, M. R. KNOX, A. V. VERSHININ and M. J. AMBROSE, 1998 Polymorphism of insertion sites of Ty1-*copia* class retrotransposons and its use for linkage and diversity analysis in pea. *Mol. Gen. Genet.* **260**: 9–19.
- FLAVELL, A. J., E. DUNBAR, R. ANDERSON, S. R. PEARCE, R. HARTLEY *et al.*, 1992 Ty1-*copia* group retrotransposons are ubiquitous and heterogeneous in higher plants. *Nucleic Acids Res.* **20**: 3639–3644.
- FRAY, R. G., and D. GRIERSON, 1993 Identification and genetic analysis of normal and mutant phytoene synthase genes of tomato by sequencing, complementation and co-suppression. *Plant Mol. Biol.* **22**: 589–602.

- FRIANT, S., T. HEYMAN, M. L. WILHELM and F. X. WILHELM, 1996 Extended interactions between the primer tRNAi (Met) and genomic RNA of the yeast Ty1 retrotransposon. *Nucleic Acids Res.* **24**: 441–449.
- HEDGES, S. B., 2002 The origin and evolution of model organisms. *Nat. Rev. Genet.* **3**: 838–849.
- JIANG, J., B. S. GILL, G. L. WANG, P. C. RONALD and D. C. WARD, 1995 Metaphase and interphase fluorescence *in situ* hybridization mapping of the rice genome with bacterial artificial chromosomes. *Proc. Natl. Acad. Sci. USA* **92**: 4487–4491.
- JIN, Y. K., and J. L. BENNETZEN, 1994 Integration and nonrandom mutation of a plasma membrane proton ATPase gene fragment within the Bsl retroelement of maize. *Plant Cell* **6**: 1177–1186.
- JORDAN, I. K., and J. F. McDONALD, 1998 Evolution of the copia retrotransposon in the *Drosophila melanogaster* species subgroup. *Mol. Biol. Evol.* **15**: 1160–1171.
- JORDAN, I. K., L. V. MATYUNINA and J. F. McDONALD, 1999 Evidence for the recent horizontal transfer of long terminal repeat retrotransposon. *Proc. Natl. Acad. Sci. USA* **96**: 12621–12625.
- KUMAR, A., and J. L. BENNETZEN, 1999 Plant retrotransposons. *Annu. Rev. Genet.* **33**: 479–532.
- KUMAR, A., S. R. PEARCE, K. MCLEAN, G. HARRISON, J. S. HESLOP-HARRISON *et al.*, 1997 The Ty1-copia group of retrotransposons in plants: genomic organisation, evolution, and use as molecular markers. *Genetica* **100**: 205–217.
- KUMAR, S., K. TAMURA and M. NEI, 2004 MEGA3: integrated software for molecular evolutionary genetics analysis and sequence alignment. *Brief. Bioinformatics* **5**: 150–163.
- LI, L., X. CHENG and H.-Q. LING, 2004 Isolation and characterization of Fe(III)-chelate reductase gene *LeFRO1* in tomato. *Plant Mol. Biol.* **54**: 125–136.
- LING, H.-Q., A. PICH, G. SCHOLZ and M. W. GANAL, 1996 Genetic analysis of two tomato mutants affected in the regulation of iron metabolism. *Mol. Gen. Genet.* **252**: 87–92.
- LING, H.-Q., P. BAUER, Z. BERCZKY, B. KELLER and M. GANAL, 2002 The tomato *fer* gene encoding a bHLH protein controls iron-uptake responses in roots. *Proc. Natl. Acad. Sci. USA* **99**: 13938–13943.
- LONG, M., S. J. DE SOUZA, C. ROSENBERG and W. GILBERT, 1996 Exon shuffling and the origin of the mitochondrial targeting function in plant cytochrome c1 precursor. *Proc. Natl. Acad. Sci. USA* **93**: 7727–7731.
- LONG, M., E. BETRAN, K. THORNTON and W. WANG, 2003 The origin of new genes: glimpses from the young and old. *Nat. Rev. Genet.* **4**: 865–875.
- LOWE, T. M., and S. R. EDDY, 1997 tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Res.* **25**: 955–964.
- MA, J., K. M. DEVOS and J. L. BENNETZEN, 2004 Analyses of LTR-retrotransposon structures reveal recent and rapid genomic DNA loss in rice. *Genome Res.* **14**: 860–869.
- MARCHLER-BAUER, A., J. B. ANDERSON, C. DEWEESE-SCOTT, N. D. FEDOROVA, L. Y. GEER *et al.*, 2003 CDD: a curated Entrez database of conserved domain alignments. *Nucleic Acids Res.* **31**: 383–387.
- NEUMANN, P., D. POZARKOVA and J. MACAS, 2003 Highly abundant pea LTR retrotransposon *Ogre* is constitutively transcribed and partially spliced. *Plant Mol. Biol.* **53**: 399–410.
- NUEZ, F., J. PROHENS and J. M. BLANCA, 2004 Relationships, origin, and diversity of Galapagos tomatoes: implications for the conservation of natural populations. *Am. J. Bot.* **91**: 86–99.
- OCHMAN, H., J. G. LAWRENCE and E. A. GROISMAN, 2000 Lateral gene transfer and the nature of bacterial innovation. *Nature* **405**: 299–304.
- PRICE, H. L., and A. W. DRINKARD, 1908 Inheritance in tomato hybrids. *Virg. Agric. Exp. Stn. Bull.* **177**: 17–53.
- PURUGGANAN, M. D., and S. R. WESSLER, 1995 Transposon signatures: species-specific molecular markers that utilize a class of multiple-copy nuclear DNA. *Mol. Ecol.* **4**: 265–269.
- RICO-CABANAS, L., and J. A. MARTINEZ-IZQUIERDO, 2007 *CIRE1*, a novel transcriptionally active Ty1-copia retrotransposon from *Citrus sinensis*. *Mol. Genet. Genomics* **277**: 365–377.
- ROBERTSON, H. M., F. N. SOTO-ADAMES, K. K. O. R. WALDEN, M. P. AYANCINI and D. J. LAMPE, 1998 The *mariner* transposons of animals: horizontally jumping genes, pp. 268–284 in *Horizontal Gene Transfer*, edited by M. SYVANEN and C. I. KADO. Chapman & Hall, London.
- ROULIN, A., B. PIEGU, R. A. WING and O. PANAUD, 2008 Evidence of multiple horizontal transfers of the long terminal repeat retrotransposon *RIRE1* within the genus *Oryza*. *Plant J.* **53**: 950–959.
- SANMIGUEL, P., A. TIKHONOV, Y. K. JIN, N. MOTCHOULSKAIA, D. ZAKHAROV *et al.*, 1996 Nested retrotransposons in the intergenic regions of the maize genome. *Science* **274**: 765–768.
- SILVA, J. C., and M. G. KIDWELL, 2000 Horizontal transfer and selection in the evolution of *Pelemets*. *Mol. Biol. Evol.* **17**: 1542–1557.
- SILVA, J. C., E. L. LORETO and J. B. CLARK, 2004 Factors that affect the horizontal transfer of transposable elements. *Curr. Issues Mol. Biol.* **6**: 57–71.
- SONNHAMMER, E. L., and R. DURBIN, 1995 A dot-matrix program with dynamic threshold control suited for genomic DNA and protein sequence analysis. *Gene* **167**: GC1–GC10.
- TANKSLEY, S. D., M. W. GANAL, J. P. PRINCE, M. C. DE VICENTE, M. W. BONIERBALE *et al.*, 1992 High density molecular linkage maps of the tomato and potato genomes. *Genetics* **132**: 1141–1160.
- VAN DER HOEVEN, R., C. RONNING, J. GIOVANNONI, G. MARTIN and S. TANKSLEY, 2002 Deductions about the number, organization, and evolution of genes in the tomato genome based on analysis of a large expressed sequence tag collection and selective genomic sequencing. *Plant Cell* **14**: 1441–1456.
- WANG, W., H. ZHENG, C. FAN, J. LI, J. SHI *et al.*, 2006 High rate of chimeric gene origination by retroposition in plant genomes. *Plant Cell* **18**: 1791–1802.
- WAUGH, R., K. MCLEAN, A. J. FLAVELL, S. R. PEARCE, A. KUMAR *et al.*, 1997 Genetic distribution of *Bare-1*-like retrotransposable elements in the barley genome revealed by sequence-specific amplification polymorphisms (SSAP). *Mol. Gen. Genet.* **253**: 687–694.
- WHITE, S. E., and J. F. DOEBLEY, 1999 The molecular evolution of *terminal ear1*, a regulatory gene in the genus *Zea*. *Genetics* **153**: 1455–1462.
- WICKER, T., and B. KELLER, 2007 Genome-wide comparative analysis of copia retrotransposons in Triticeae, rice, and Arabidopsis reveals conserved ancient evolutionary lineages and distinct dynamics of individual copia families. *Genome Res.* **17**: 1072–1081.
- WU, H. K., 1967 Note on preparing of pachytene chromosomes by double mordants. *Sci. Agric.* **15**: 40–44.
- XIAO, H., N. JIANG, E. SCHAFFNER, E. J. STOCKINGER and E. VAN DER KNAAP, 2008 A retrotransposon-mediated gene duplication underlies morphological variation of tomato fruit. *Science* **319**: 1527–1530.