Chapter 15
Rider Transposon Insertion and Phenotypic Change in Tomato

Ning Jiang, Sofia Visa, Shan Wu, and Esther van der Knaap

Abstract The *Rider* retrotransposon is ubiquitous in the tomato genome and is likely an autonomous element that still transposes to date. The majority of approximately 2,000 copies of *Rider* are located near genes. Phenotypes associated with *Rider* insertion are diverse and often the result of knock out of the underlying genes. One unusual *Rider*-mediated phenotype resulted from a gene duplication event. By means of read-through transcription, *Rider* copied part of the surrounding sequence to another location in the genome, leading to high expression of one of the transposed genes, *SUN*, resulting in an elongated fruit shape. Transcription studies demonstrated that *Rider* is expressed to levels comparable to the expression of other tomato genes and that control of transposition may be regulated by antisense transcription. Taken together, *Rider* is a unique retrotransposon that may have played important roles in the evolution of tomato and its closest relatives.

Keywords LTR Copia • Phenotype • Rider • Tomato • Transcription

Abbreviations

ATP	Adenosine triphosphate
bHLH	Basic helix–loop–helix
BL	Blind
C	Cut leaf or potato leaf mutation

N. Jiang
Department of Horticulture, Michigan State University, East Lansing, MI 48824, USA

S. Visa
Department of Mathematics and Computer Science, College of Wooster, Wooster, OH 44691, USA

S. Wu • E. van der Knaap (✉)
Department of Horticulture and Crop Science, The Ohio State University, Wooster, OH 44691, USA
e-mail: Vanderknaap.1@osu.edu

M.-A. Grandbastien and J.M. Casacuberta (eds.), *Plant Transposable Elements*, Topics in Current Genetics 24, DOI 10.1007/978-3-642-31842-9_15,

CP	Coat protein
DNA	Deoxyribonucleic acid
EST	Expressed sequence tag
FER	Iron inefficient mutant
INT	Integrase
LTR	Long terminal repeat
Mb	Mega base pair
MITE	Miniature inverted repeat transposable element
mRNA	messenger RNA
MULE	Mutator-like element
MYA	Million years ago
MYB	Myeloblastosis transcription factor
PBS	Primer binding site
PPT	Polypurine tract
PR	Protease
PSY1	Phytoene synthase 1
R	Red or yellow flesh mutation
RAX1	Regulator of axillary meristem 1
RH	RNase H
RISC	RNA-induced silencing complex
RNA	Ribonucleic acid
TE	Transposable element
TIR	Terminal inverted repeat
TSD	Target site duplication

15.1 The Abundance of TEs in Genomes and the Phenotypic Consequences of their Insertions

Transposable elements (TEs), DNA fragments capable of replication and movement, are major components of eukaryotic genomes. Depending on the timing of their transposition activity, they may display different insertion sites among closely related genomes and hence contribute to genome diversity. TEs are divided into two classes. Class I elements or RNA elements (retrotransposons) use the element-encoded mRNA as the transposition intermediate. These transposons are either flanked by a long terminal repeat (LTR) or lack terminal repeat sequences (non-LTR transposons). Class II elements or DNA transposons are often characterized by the terminal inverted repeats (TIRs) and transposition through a DNA intermediate. Autonomous TEs encode a transposase and other proteins required for transposition, while nonautonomous elements lack functional transposition proteins and rely on the cognate autonomous TEs for their transposition. In plants, LTR retrotransposons are very abundant and are largely responsible for the genome size expansion in grass species (Bennetzen 1996). This is also the case for species in the *Solanaceae* family that includes tomato (*Solanum lycopersicum*), potato (*S. tuberosum*),

pepper (*Capsicum* spp.), eggplant (*S. melongena*), petunia (*Petunia* spp.), and tobacco (*Nicotiana* spp.). The different genome size that ranges from 844 [potato, (Consortium 2011)] to 4,500 Mb (*Nicotiana tabacum*) is largely attributed to differences in the number of LTR elements, some of which are found in the euchromatic parts of the genome (Park et al. 2011a, b). Reduction in genome size also occurs by unequal recombination between the two LTRs of a single element (Ma et al. 2004). This often leads to the deletion of the internal region and one of the LTRs resulting in the formation of a "solo" LTR.

Transposons are mostly known for the disruption of genes when they insert into or very close to genes. However, they are also known to duplicate and mobilize gene sequences. Recent studies indicate most major types of TEs are capable of duplicating and amplifying genes or gene fragments (Kazazian 2004; Bennetzen 2005; Feschotte and Pritham 2007; Schnable et al. 2009). For example, the maize *Bs1* LTR retrotransposon carries part of a plasma membrane proton-translocating ATPase gene without its intron sequences (Bureau et al. 1994; Jin and Bennetzen 1994). Subsequently, it was shown that this chimeric element was transcribed and translated in early ear development and might have a function in the reproductive pathway (Elrouby and Bureau 2010). In rice, over a thousand genes that duplicated through retrotransposition (retrogenes) have been identified, and many recruited new exons from flanking regions, resulting in the formation of chimeric genes (Wang et al. 2006a). Similarly, there are thousands of *Mutator*-like elements (MULE) that carry genes or gene fragments in the rice genome (Jiang et al. 2004; Juretic et al. 2005). Due to the ability to duplicate genes or gene fragment, transposons themselves may represent the structural variation among species or individuals in the population. For example, there are thousands of *Helitron*s carrying genes in maize (Du et al. 2009; Yang and Bennetzen 2009), and they contribute significantly to many fragments that are not shared among different maize cultivars at the orthologous position (Fu and Dooner 2002; Morgante et al. 2005).

Despite the abundance of transposons in the tomato genome, few are known to result in an altered phenotype. In this chapter, we summarize the findings of what is known about *Rider*, a high copy *Copia* element found in tomato and its closest wild relatives. The element was first described as the cause of the elongated fruit shape at the locus *sun* and its ability to duplicate genes from one chromosome to another (Xiao et al. 2008). In addition, there are many unusual features of *Rider* that warrant further investigations as will be demonstrated below.

15.2 Features Associated with *Rider*

15.2.1 The Structure of the Rider *Element*

The structure of *Rider* element resembles that of a typical *Copia*-like element from many perspectives. The element is 4,867 in length with two identical LTRs on each terminus (Fig. 15.1). The LTR of the *Rider* element at *SUN* is 398 bp in length and

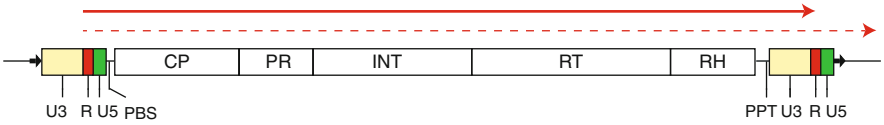

Fig. 15.1 The structure of *Rider*. *Color boxes* indicate distinct regions in LTR (U3, R and U5). Coding regions are indicated as *white boxes*. The genes within *Rider* are shown as *white boxes* and encode capsid-like proteins (CP), protease (PR), integrase (INT), reverse transcriptase (RT), and RNase-H (RH). Other sequence features are primer binding site (PBS), polypurine tract (PPT). *Black arrows* flanking the LTRs indicate target site duplication (TSD). For *Rider* elements, TSD are 5 bp. *Red solid arrow* represents a normal transcript from *Rider*, while the *dashed arrow* exemplifies a read-through transcript

includes the three classical LTR domains called U3, R, and U5. U3 region contains the promoter of the element, and its size is highly variable among individual *Rider* elements (Jiang et al. 2009). Sequences in R and U5 are responsible for the termination of transcription of the element, and they are well conserved among most individual elements (also see below). The internal region of *Rider* is 4,071 bp and encodes a single polyprotein of 1,307 amino acids, accounting for 96% of the internal region. The polyprotein contains all typical proteins or domains that a *Copia*-like element encodes, including capsid-like protein, protease, integrase, reverse transcriptase, and RNase H (Fig. 15.1) (Kumar and Bennetzen 1999). The internal region also contains the cis-elements required for transposition, such as the primer binding site and polypurine track (Lewin 2008). Thus, *Rider* is likely to be an autonomous *Copia*-like element.

15.2.2 The Timing of Rider *Amplification*

Database searches and DNA blots using the LTR as probe indicate that *Rider* element is present in all *Solanum* section *Lycopersicon* species tested and absent from related species such as potato, tobacco, and coffee (Cheng et al. 2009; Jiang et al. 2009). Therefore, it appears that the initial amplification of *Rider* occurred prior to the speciation of *Lycopersicon* section species and after the divergence of tomato and potato, which is estimated to be between 5.1 and 7.3 MYA (Wang et al. 2008). Among the section *Lycopersicon* species, variation of copy number was observed. For example, the copy number of *Rider* appears to be lower in the genomes of *S. habrochaites* and *S. chilense* compared to other species (Cheng et al. 2009; Jiang et al. 2009). The tomato genome harbors about 2,000 copies of *Rider* based on partial genome sequence surveys (Jiang et al. 2009). Two-thirds of the intact *Rider* elements inserted after the divergence of *S. lycopersicum* and *S. pimpinellifolium*, which occurred about 1.3 MYA. This finding suggests that the majority of *Rider* elements arose well after the speciation in the section *Lycopersicon* (Jiang et al. 2009). Moreover, insertion polymorphism of *Rider* and transcript accumulation were detected among different tomato cultivars

(Cheng et al. 2009; Jiang et al. 2009). Due to the high insertion polymorphism among tomato species, *Rider* would be useful as a tool for studying the phylogenic relationship in this important group.

15.2.3 The Origin of the Rider *Element*

The origin of *Rider* is mysterious. The presence of TEs in a certain genome can be either due to vertical transmission from ancestral genomes or horizontal transfer from an unrelated species. As mentioned above, *Rider* is absent from potato, tobacco, and coffee. Meanwhile, two individual LTR elements in *Arabidopsis*, named *Rider-like* 1 and *Rider-like* 2, have moderate nucleotide similarity (~75%) with *Rider* in the internal region and part of the LTR sequence (Cheng et al. 2009). For this reason, it was proposed that *Rider* was introduced into the tomato genome 1–6 MYA from *Arabidopsis* or a relative of *Arabidopsis* (Cheng et al. 2009). While the similarity between *Rider-like* elements and *Rider* is unusually high given the genetic distance between *Arabidopsis* and tomato, there is not sufficient evidence to support an unambiguous case of direct transfer between the two species in the proposed timeframe. Elements highly similar to *Rider-like* elements are not present in genomes of species related to *Arabidopsis*, such as *A. lyrata* and *B. oleracea* ((Cheng et al. 2009), Jiang, unpublished data). As a result, the ultimate donor or ancestor of *Rider* is unclear if it indeed resulted from horizontal transfer from one to the other species.

An equally plausible explanation for the occurrence of *Rider* and *Rider*-like elements in two distant genomes is that *Rider* is inherited from the ancestral genome of tomato and lost from related species. This is because most TE families experience a life cycle of "birth–burst–extinction" (Hartl et al. 1997). Once a TE family is no longer transpositionally active, mutations and deletions accumulate and the particular family will eventually disappear from the genome. According this scenario, loss of TEs from a genome is a common event and only a small subset of TEs can achieve long-term success. Due to the fact that *Rider* is a compact element without obviously nonessential sequences, the conservation between *Rider* and *Rider-like* elements could be due to functional constraints. Consequently, the origin of *Rider* is still an open question. The clarification of this issue awaits the availability of more genomic sequences in *Brassica* and *Solanaceae*, and other plant species.

15.3 Distribution and Targeting Preference of *Rider*

Plant genomes harbor numerous types of transposons, and different transposons have distinct niches. The distribution pattern of any transposons, including LTR elements, is the consequence of target specificity and selection against deleterious insertions or selection for favorable insertions (Pereira 2004; Peterson-Burch et al. 2004).

Many high copy number LTR elements are nested in the intergenic or heterochromatic regions (SanMiguel et al. 1996; Ananiev et al. 1998; Jiang et al. 2002). In contrast, low copy number LTR elements, such as *Tpv2* elements (40 copies) in common bean (*Phaseolus vulgaris*) and *Tos17* (a few copies in natural populations) in rice (*Oryza sativa*), are frequently found in genic regions (Garber et al. 1999; Miyao et al. 2003). Given the fact that *Tos17* can amplify rapidly under artificial conditions (Hirochika et al. 1996), its low copy number in natural populations suggests that the preference for genic regions may result in deleterious effect on the host organism, which prevents the element from further amplifications. The only known exception is the *Tnt1* element from tobacco, which has a relatively high copy number (a few hundred copies), yet is located in genic regions (Grandbastien et al. 1989; Le et al. 2007). Nevertheless, *Tnt1*-related elements are only present in a few dozens in tomato and those are mostly mapped to pericentromeric regions (Tam et al. 2007), suggesting host environment may have important influence on amplification and distribution of LTR elements.

Unlike any of the known LTR elements, *Rider* does not appear to be concentrated in certain regions of the genome (Cheng et al. 2009). Moreover, about half of the *Rider* elements are located within 1 kb of a gene. This ratio is much higher than that of another high copy number tomato LTR element *Jinling*, for which only 20% of the elements are within the same distance to a gene (Jiang et al. 2009). This can be explained by the difference in their chromosomal distribution patterns since most *Jinling* elements are located in heterochromation regions where the gene density is low (Wang et al. 2006b). In contrast, *Rider* elements are located in both heterochromatic and euchromatic regions so they are more likely surrounded by genes.

Despite its high copy number and frequent associations with genes, *Rider* does not seem to disrupt genes at a high level that would render the tomato genome unstable. This could be due to the regulation of its expression (see below) and to its insertion preference. *Rider* appears to insert into AT-rich sequence (Jiang et al. 2009). Since coding regions are usually more GC-rich than noncoding regions (Salinas et al. 1988; Mizuno and Kanehisa 1994), such a preference allows *Rider* elements to select noncoding regions as their targets and minimize possible deleterious effects. In this case, the amplification of *Rider* is largely silent despite the fact that many elements are close to genes. Meanwhile, being located in the genic regions may favor the element amplification since the element is more accessible to the transcription machinery. This might partially explain the success of *Rider* in the tomato genome.

15.4 Rider Expression, Read-Through Transcription and its Correlation with Mutations in LTR

Based on Northern blot and RT-PCR experiments as well as database searches, *Rider* is constitutively expressed in tomato (Cheng et al. 2009; Jiang et al. 2009). Transcript sizes suggest that most *Rider* RNA is intact and has the potential to

transpose to new positions. Mining through mRNA seq data sets also showed that *Rider* is expressed in certain tissues at a level comparable to the tomato fruit shape gene *OVATE*. *SUN* and *R* (the latter corresponding a phytoene synthase gene, see below) are expressed higher than *Rider* while *DEFL2* (encoding a defensin protein, see below) is expressed the highest in the tissues examined (Table 15.1). Interestingly, while only sense expression of the genes *SUN*, *OVATE*, *R,* and *DEFL2* is found, *Rider* appears to be expressed equally in both sense and antisense direction (Table 15.1), raising the interesting question of whether the regulation of transposition is mediated in part by posttranscriptional silencing. Further examination of the position of the mRNA seq reads relative to *Rider* revealed that the reads are evenly distributed along the transposon in both directions (Fig. 15.2). Due to the finding that intact *Rider* elements outnumber truncated elements by 3.5 to 1, this suggests that transcription in the sense and antisense direction are derived from intact elements. However, spurious expression from exogenous promoters into truncated *Rider* elements cannot be excluded either. Regardless, double stranded RNAs are commonly resulting in rapid mRNA degradation via the RNA-induced silencing complex (RISC). Therefore, the potential gene silencing of *Rider* might explain the results from Northern blots that showed smears instead of one distinct band (Jiang et al. 2009). This finding is also consistent with the observation that the insertion polymorphism of *Rider* among tomato cultivars is relatively low compared to that among *Solanum* subsection *Lycopersicon* species (Jiang et al. 2009). In other words, the high copy number of *Rider* is likely due to its high transposition activity in recent past, which may have declined due to potential silencing arising from the abundance of elements.

A low number of ESTs were found to be chimeric between *Rider* LTR and an unrelated sequence (Cheng et al. 2009; Jiang et al. 2009). These chimeric elements can be explained by artifacts in the construction of the library for EST. Alternatively, these aberrant RNAs could also lead to gene silencing in cases where the chimeric part exhibits high sequence similarity to an endogenous gene. The finding of chimeric EST reads could also be the result of read-through transcription. Normally, *Rider* transcription starts in the R region of the 5′ LTR and ends in the R region of the 3′ LTR. Read-through transcription would extend past the R region into the U5 and neighboring genome region. Indeed, read-through transcription of *Rider* is found in all the tissues examined (Jiang et al. 2009).

Read-through transcription is at the heart of the *SUN* duplication as will be discussed in detail below. The *Rider* element that created the locus carried a mutation in one of the two "TTGT" sequences required for transcript termination (Jiang et al. 2009). The sequencing of read-through transcripts over the region that is required for termination indeed showed that the majority of the transcripts carried the mutation in the LTR found in the *Rider* element at the *sun* locus. These findings strongly suggest that read-through transcripts are indeed associated with *Rider* elements and are more prevalent when the LTR carries the "TTAT" mutation in one of the "TTGT" copies in the U5 region.

Table 15.1 Expression of *Rider* and four genes in tomato flower buds

Sample	Strand	RIDER	RIDER RPKM	SUN	SUN RPKM	OVATE	OVATE RPKM	DEFL2	DEFL2 RPKM	R	R RPKM	Number of illumina reads aligned to the genome (millions)
Replicate 1	Antisense	183	2.92	3	0.11	0	0	0	0	0	0	
	Sense	225	3.59	1,134	39.83	84	5.73	1,053	165.64	383	18.23	
	Total	408	6.52	1,137	39.94	84	5.73	1,053	165.64	383	18.23	13.82
Replicate 2	Antisense	268	3.83	1	0.03	0	0	0	0	0	0	
	Sense	342	4.89	1,209	38.01	87	5.32	1,196	168.39	959	40.86	
	Total	610	8.72	1,210	38.04	87	5.32	1,196	168.39	959	40.86	15.44
Replicate 3	Antisense	382	5	2	0.06	0	0	0	0	0	0	
	Sense	278	3.64	1,554	44.74	141	7.89	710	91.55	477	18.61	
	Total	660	8.64	1,556	44.8	141	7.89	710	91.55	477	18.61	16.86
Gene length (Kb)		4.53		2.06		1.06		0.46		1.52		

Small flower bud tissues from the tomato accession SA2 (LA1589 near isogenic line carrying the *SUN* gene duplication; Xiao et al. 2008) were harvested and RNA was extracted. Directional libraries were constructed according to Zhong et al. (2011). Reads were aligned to the five sequences using Tophat and allowing one mismatch. The first column for each sequence shows the number of raw reads. RPKM = reads per kb per million reads. Accession numbers: Rider (SGN-U569744), SUN (EU491503), OVATE (AY140893), DEFL2 (Solyc07g007750), R (Solyc03g031860). Due to the high similarity of *Rider* elements, the aligned reads cannot be assigned to specific elements in the tomato genome

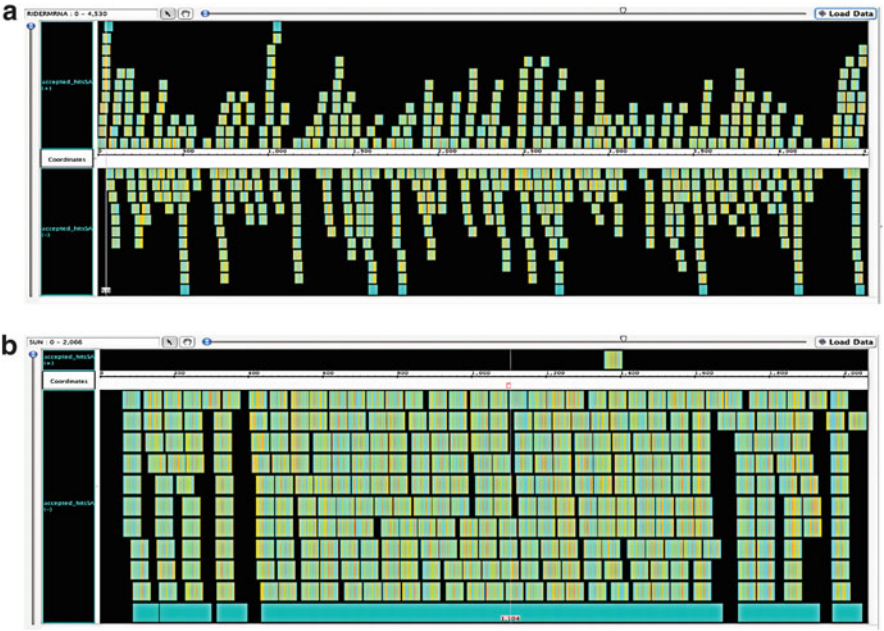

Fig. 15.2 Alignment of the mRNA seq reads to (**a**) *Rider* and (**b**) *SUN*. The libraries were constructed strand-specifically such that only the first-strand cDNA will yield reads (Zhong et al. 2011). The SAM files generated by Tophat alignment were visualized using the IGB viewer (http://bioviz.org/igb/faq.shtml). *Rider* reads are found in both directions along the transposon while *SUN* reads are only found for the sense strand (− strand in the viewer), with the exception of one read. Results are from replicate 2 in Table 15.1. The *solid green bar* on the bottom indicates that there are more reads corresponding to *SUN* that are not displayed in the viewer due to space constraints

15.5 Case Studies of Phenotypic Changes Caused by *Rider* and Genomic Landscape in which the Element Inserts

As demonstrated in the previous sections, the insertion preference of *Rider* is found near genes, *Rider* is constitutively expressed in tomato tissues albeit in both directions and *Rider* read-through transcription is occurring. Also, *Rider* elements are only found in the species of the *Solanum* section *Lycopersicon* and not in other Solanaceous relatives such as potato and tobacco. In addition, *Rider* has been shown to be involved in phenotypic changes that are found in the *Lycopersicon* section of the Solanum genus, including those that impact domestication-related phenotypes as well as spontaneously arising mutations.

15.5.1 Rider *and Fruit Shape*

One of the most striking examples of phenotypic change mediated by *Rider* transposition is found at the fruit shape locus *sun* located on chromosome 7 (Xiao et al. 2008).

The locus resulted from a *Rider* transposition in which nearly 20 kb of the neighboring genome was included in the event. Based on sequence comparisons, the transposition and resulting genomic duplication was deduced to have happened as follows. Read-through transcription of the *Rider* element on chromosome 10 found at position 60,134,479–60,139,738 (http://www.solgenomics.net unigene SGN-U569744) into the neighboring genes, followed by a template switch in the first intron of a SDL1-like gene to downstream of an IQ domain-containing gene found at position 60,140,568–60,142,797. Transcription continued until the first LTR of *Rider* (Xiao et al. 2008). This giant retroelement, that includes *Rider* and nearby genome sequence, transposed into the intron of *DEFL1* located on chromosome 7 at position 2,394,467–2,396,320 (Solyc07g007760) (Jiang et al. 2009). The IQ domain-containing gene that originated from chromosome 10 is located in a new genome environment leading to high expression in the fruit resulting in an elongated fruit shape (Xiao et al. 2008). Thus the IQ domain containing gene was renamed *SUN*. The *Rider* insertion knocked out the expression of *DEFL1* (Solyc07g007760) and reduced the expression of the neighboring *DEFL2* gene (Solyc07g007750) by at least fivefold (unpublished mRNA seq data). Further studies have shown that the transposition of *Rider* and duplication of *SUN* was most likely a post-domestication event originating in Europe in the last 200–500 years (Rodriguez et al. 2011). Varieties carrying the *SUN* duplication result in fruit with an almost pepper-like or oxheart shape, which are typically found in heirloom tomatoes (Fig. 15.3). The genome environment of the ancestral locus on chromosome 10 showed no class I transposons except for *Rider*, but instead a high number of class II DNA transposons. At the *sun* locus, the number of class II transposons was higher than found on the ancestral locus (Jiang et al. 2009).

15.5.2 Rider *and Iron Deficiency*

The chlorotic tomato mutant *fer* was a spontaneous mutant identified in the 1960s (Brown et al. 1971). The mutant plant exhibits defects in all the typical responses to iron deficiency and uptake of Fe^{3+} (Brown et al. 1971; Ling et al. 1996). Although located in the pericentromeric region of chromosome 6, which might exhibit reduced recombination rates, the *FER* gene was identified by positional cloning and found to encode a bHLH protein involved in the transcriptional regulation of plant iron nutrition (Ling et al. 2002; Brumbarova and Bauer 2005; Guyot et al. 2005). The gene is located on chromosome 6 at position 31,549,026–31,547,113 (Solyc06g051550). The mutation in tomato *FER* was due to a spontaneous insertion of *Rider* in the first exon resulting in disruption of the gene (Ling et al. 2002; Cheng et al. 2009). The *fer Rider* element is 100% identical, including the LTRs, to the *Rider* element found at the *sun* locus and the ancestral locus on chromosome 10 (Cheng et al. 2009). A high level of transposable elements, including class I, class II and unclassified repeats, are found at the *fer* locus demonstrating a highly diverse TE landscape in the pericentromeric region of chromosome 6. The *fer* locus also features a relatively low density of genes of 19.8 kb per gene (Guyot et al. 2005).

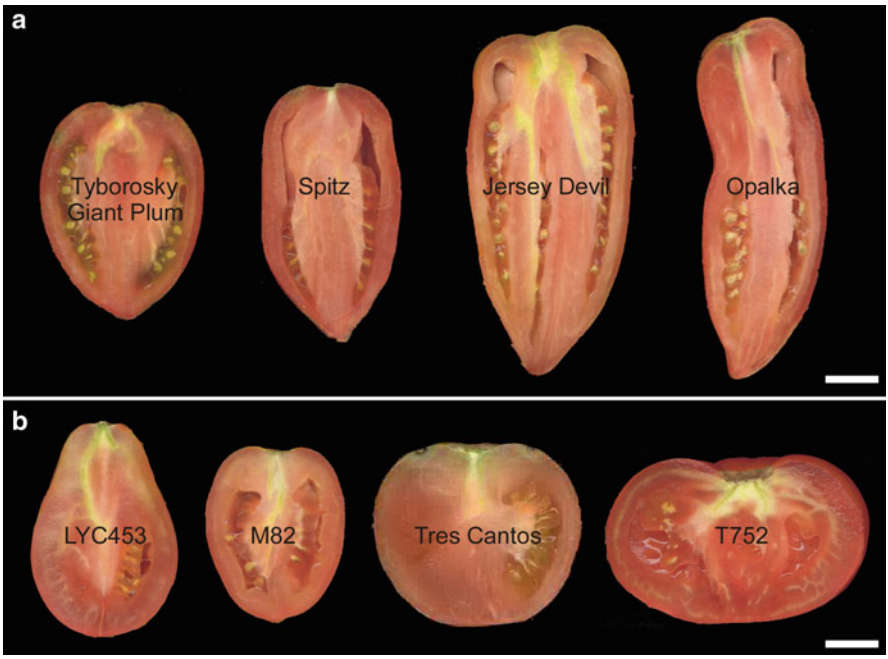

Fig. 15.3 Tomato fruit shape affected by *Rider*. (**a**) Varieties with the *SUN* gene resulting from *Rider* transposition and gene duplication. (**b**) Varieties without the *SUN* gene duplication. The variety names are written in each fruit (Rodriguez et al. 2011). Note the characteristically long fruit and pointed shape as a result of *SUN*. Pear-shaped fruit (LYC453 in B) is controlled by *OVATE*. *Bar* corresponds to 2 cm

This is in contrast to the *sun* locus and the ancestral locus on chromosome 10, where gene density approached that of what is typically found in euchromatin in the range of 5–7 kb per gene (Jiang et al. 2009). Other than *Rider*, the LTR transposons found at the *fer* locus are neither active nor autonomous as they have accumulated numerous mutations (Guyot et al. 2005).

15.5.3 Rider *and Fruit Color*

The yellow flesh mutation in tomato confers a yellow instead of the wild type red fruit and the locus is named "r" (Price and Drinkard 1908). The underlying gene is *phytoene synthase 1* (*PSY1*) that encodes the first enzyme in the carotenoid biosynthesis pathway. Initially, the gene was identified in a screen for ripening-induced genes (Bartley et al. 1992; Fray and Grierson 1993). The cDNA cloning and sequencing of the two allelic versions of the yellow flesh mutant alleles, *r* and r^y, showed that the older allele, *r*, was due to an insertion of a repetitive element (Fray and Grierson 1993). Sequence comparisons of the inserted fragment of

328 nucleotides showed that it corresponded to the LTR of *Rider* with 96% identity to the element found at the *sun* locus. *PSY1* is found on chromosome 3 at position 8,606,368–8,610,361 (Solyc03g031860). A detailed analysis of the genome structure at the *r* locus has not been conducted. However, the r^y allele appears to be the result of a short deletion because the 3′ end of the cDNA sequence of the mutant *psy1* gene corresponds to a region approximately 4.5 kb downstream of *PSY1* comprising the first exon of an Acyl-CoA synthase gene (Solyc03g031870). This finding suggests that the *r* locus may have experienced other types of rearrangement unrelated to *Rider* transposition.

15.5.4 Rider *and Leaf Complexity*

The last and most recently reported example of a phenotypic change mediated by *Rider* transposition is exemplified by the gene underlying the "potato leaf" mutation in tomato. The locus is called *C*, for cut leaf. Tomato features complex leaves comprised of terminal and lateral leaflets that are often serrated at the margins. The potato leaf represents an old tomato mutation resulting in reduced leaf complexity and smooth leaf blade margins (Price and Drinkard 1908; Busch et al. 2011). The underlying gene is a member of the R2R3 MYB transcription factor family that is evolutionarily very closely related to the tomato *BLIND* (*BL*) gene regulating shoot branching. *C* (Solyc06g074910) maps to chromosome 6 at position 42,804,036–42,806,196. *C* has acquired a new but related function compared to *BL* and both correspond to *RAX1* in Arabidopsis regulating shoot branching (Busch et al. 2011). *Rider* inserted near the 3′ end of *C* disrupting the coding region resulting in a null mutation. The *Rider* element found at *c* is identical in sequence to the element found at *sun* (Busch et al. 2011). Except for the *Rider* insertion allele which is spontaneous, most of the other reported *c* alleles were derived from mutagenesis screens (Busch et al. 2011). Of these induced mutations, two resulted from a deletion event of 286 bp and 40.6 kb, respectively. Although a detailed genome analysis of the locus has not been conducted, the *c* locus also appears prone to genome rearrangements in addition to transposon insertions.

15.6 Concluding Remarks

Transposable elements achieve their success through different strategies. Some elements, such as *Jinling* in tomato, are preferentially located in the pericentromeric heterochromatin, which is the "safe haven" for insertion. Other elements, such as the miniature inverted repeat transposable element (MITE) *mPing* in rice, are preferentially located in genic regions. Nevertheless, the impact of MITE insertion is often subtle due to their small size (usually less than 500 bp) as well as avoidance of insertion into coding region (Naito et al. 2009). Moreover, *mPing* harbors regulatory

motifs that enable the adjacent genes to become stress inducible (Naito et al. 2009). In other words, a successful transposable element must either have minimal detrimental impact or bring about favorable mutations for the host genome, especially when the element is capable of transposition. From this point of view, *Rider* has developed many features for its success despite its relatively large size. First of all, *Rider* elements have been active in transposition since it amplified to thousands of copies in just a few million years. The most recent known transposition occurred in the 1960s with the creation of the *fer* locus (Cheng et al. 2009). Second, it targets all chromosomal regions but appears to avoid inserting into coding regions by selective insertions into AT-rich regions. Third, the transposition activity of *Rider* is likely regulated by antisense transcription of the element, thereby limiting the extent of transposition per generation. Finally, *Rider* creates read-through transcripts which may allow the duplication of flanking sequences including genes. The duplication of genes may create novel phenotypes that are favored by selection. Taken together, *Rider* is a unique retrotransposon that has been successfully amplified in the genome of tomato and may have played important roles in the evolution of tomato and its closest relatives.

Acknowledgments Funding in the Jiang laboratory is provided by National Science Foundation Molecular and Cellular Biosciences grant number 1121650. Funding in the van der Knaap laboratory is provided by National Science Foundation Integrative Organismal Systems grant number 0922661.

References

Ananiev EV, Phillips RL, Rines HW (1998) Chromosome-specific molecular organization of maize (*Zea mays* L.) centromeric regions. Proc Natl Acad Sci USA 95:13073–13078

Bartley GE, Viitanen PV, Bacot KO, Scolnik PA (1992) A tomato gene expressed during fruit ripening encodes an enzyme of the carotenoid biosynthesis pathway. J Biol Chem 267:5036–5039

Bennetzen JL (1996) The contributions of retroelements to plant genome organization, function and evolution. Trends Microbiol 4:347–353

Bennetzen JL (2005) Transposable elements, gene creation and genome rearrangement in flowering plants. Curr Opin Genet Dev 15:621–627

Brown JC, Chaney RL, Ambler JE (1971) A new tomato mutant inefficient in the transport of iron. Physiol Plant 25:48–53

Brumbarova T, Bauer P (2005) Iron-mediated control of the basic helix-loop-helix protein FER, a regulator of iron uptake in tomato. Plant Physiol 137:1018–1026

Bureau TE, White SE, Wessler SR (1994) Transduction of a cellular gene by a plant retroelement. Cell 77:479–480

Busch BL, Schmitz G, Rossmann S, Piron F, Ding J, Bendahmane A, Theres K (2011) Shoot branching and leaf dissection in tomato are regulated by homologous gene modules. Plant Cell 23:3595–3609

Cheng X, Zhang D, Cheng Z, Keller B, Ling H-Q (2009) A new family of Ty1-*copia*-like retrotransposon originated in the tomato genomes by a recent horizontal transfer event. Genetics 181:1183–1193

Consortium PGS (2011) Genome sequence and analysis of the tuber crop potato. Nature 475:189–195
Du C, Fefelova N, Caronna J, He L, Dooner HK (2009) The polychromatic Helitron landscape of the maize genome. Proc Natl Acad Sci USA 106:19916–19921
Elrouby N, Bureau TE (2010) *Bs1*, a new chimeric gene formed by retrotransposon-mediated exon shuffling in maize. Plant Physiol 153:1413–1424
Feschotte C, Pritham E (2007) DNA transposons and the evolution of eukaryotic genomes. Annu Rev Genet 41:331–368
Fray RG, Grierson D (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
Fu H, Dooner HK (2002) Intraspecific violation of genetic colinearity and its implications in maize. Proc Natl Acad Sci USA 99:9573–9578
Garber K, Bilic I, Pusch O, Tohme J, Bachmair A, Schweizer D, Jantsch V (1999) The *Tpv2* family of retrotransposons of *Phaseolus vulgaris*: structure, integration characteristics, and use for genotype classification. Plant Mol Biol 39:797–807
Grandbastien MA, Spielmann A, Caboche M (1989) Tnt1, a mobile retroviral-like transposable element of tobacco isolated by plant cell genetics. Nature 337:376–380
Guyot R, Cheng X, Su Y, Cheng Z, Schlagenhauf E, Keller B, Ling H-Q (2005) Complex organization and evolution of the tomato pericentromeric region at the *FER* gene locus. Plant Physiol 138:1205–1215
Hartl DL, Lozovskaya ER, Nurminsky DI, Lohe AR (1997) What restricts the activity of mariner-like transposable elements? Trends Genet 13:197–201
Hirochika H, Sugimoto K, Otsuki Y, Tsugawa H, Kanda M (1996) Retrotransposons of rice involved in mutations induced by tissue culture. Proc Natl Acad Sci USA 93:7783–7788
Jiang N, Bao Z, Temnykh S, Cheng Z, Jiang J, Wing RA, McCouch SR, Wessler SR (2002) *Dasheng*: a recently amplified nonautonomous long terminal repeat element that is a major component of pericentromeric regions in rice. Genetics 161:1293–1305
Jiang N, Bao Z, Zhang X, Eddy SR, Wessler SR (2004) Pack-MULE transposable elements mediate gene evolution in plants. Nature 431:569–573
Jiang N, Gao D, Xiao H, van der Knaap E (2009) Genome organization of the tomato *sun* locus and characterization of the unusual retrotransposon *Rider*. Plant J 60:181–193
Jin YK, Bennetzen JL (1994) Integration and nonrandom mutation of a plasma membrane proton ATPase gene fragment within the *Bs1* retroelement of maize. Plant Cell 6:1177–1186
Juretic N, Hoen DR, Huynh ML, Harrison PM, Bureau TE (2005) The evolutionary fate of MULE-mediated duplications of host gene fragments in rice. Genome Res 15:1292–1297
Kazazian HH Jr (2004) Mobile elements: drivers of genome evolution. Science 303:1626–1632
Kumar A, Bennetzen JL (1999) Plant retrotransposons. Annu Rev Genet 33:479–532
Le QH, Melayah D, Bonnivard E, Petit M, Grandbastien MA (2007) Distribution dynamics of the Tnt1 retrotransposon in tobacco. Mol Gen Genet 278:639–651
Lewin B (2008) Genes IX. Jones and Bartlett Publishers, Sudbury, MA
Ling H-Q, Pitch A, Scholz G, Ganal MW (1996) Genetic analysis of two tomato mutants affected in the regulation of iron metabolism. Mol Gen Genet 252:87–92
Ling HQ, Bauer P, Bereczky Z, Keller B, Ganal M (2002) The tomato *fer* gene encoding a bHLH protein controls iron-uptake responses in roots. Proc Natl Acad Sci USA 99:13938–13943
Ma J, Devos KM, Bennetzen JL (2004) Analyses of LTR-retrotransposon structures reveal recent and rapid genomic DNA loss in rice. Genome Res 14:860–869
Miyao A, Tanaka K, Murata K, Sawaki H, Takeda S, Abe K, Shinozuka Y, Onosato K, Hirochika H (2003) Target site specificity of the *Tos17* retrotransposon shows a preference for insertion within genes and against insertion in retrotransposon-rich regions of the genome. Plant Cell 15:1771–1780
Mizuno M, Kanehisa M (1994) Distribution profiles of GC content around the translation initiation site in different species. FEBS Lett 352:7–10

Morgante M, Brunner S, Pea G, Fengler K, Zuccolo A, Rafalski A (2005) Gene duplication and exon shuffling by helitron-like transposons generate intraspecies diversity in maize. Nat Genet 37:997–1002

Naito K, Zhang F, Tsukiyama T, Saito H, Hancock CN, Richardson AO, Okumoto Y, Tanisaka T, Wessler SR (2009) Unexpected consequences of a sudden and massive transposon amplification on rice gene expression. Nature 461:1130–1134

Park M, Jo S-H, Kwon J-K, Park J, Ahn J-H, Kim S, Lee Y-H, Yang T-J, Hur C-G, Kang B-C, Kim B-D, Choi D (2011a) Comparative analysis of pepper and tomato reveals euchromatin expansion of pepper genome caused by differential accumulation of Ty3/Gypsy-like elements. BMC Genomics 12:85

Park M, Park J, Kim S, Kwon J-K, Park HM, Bae IH, Yang T-J, Lee Y-H, Kang B-C, Choi D (2011b) Evolution of the large genome in *Capsicum annuum* occurred through accumulation of single-type long terminal repeat retrotransposons and their derivatives. Plant J 69:1018–1029

Pereira V (2004) Insertion bias and purifying selection of retrotransposons in the *Arabidopsis thaliana* genome. Genome Biol 5:R79

Peterson-Burch BD, Nettleton D, Voytas DF (2004) Genomic neighborhoods for Arabidopsis retrotransposons: a role for targeted integration in the distribution of the Metaviridae. Genome Biol 5:R78

Price HL, Drinkard AWJ (1908) Inheritance in tomato hybrids. Virginia Agr Exp Sta Bull 177:18–53

Rodriguez GR, Munos S, Anderson C, Sim SC, Michel A, Causse M, Gardener BB, Francis D, van der Knaap E (2011) Distribution of SUN, OVATE, LC, and FAS in the tomato germplasm and the relationship to fruit shape diversity. Plant Physiol 156:275–285

Salinas J, Matassi G, Montero LM, Bernardi G (1988) Compositional compartmentalization and compositional patterns in the nuclear genomes of plants. Nucl Acids Res 16:4269–4285

SanMiguel P, Tikhonov A, Jin Y-K, Motchoulskaia N, Zakharov D, Melake-Berhan A, Springer PS, Edwards KJ, Lee M, Avramova Z, Bennetzen JL (1996) Nested retrotransposons in the intergenic regions of the maize genome. Science 274:765–768

Schnable PS, Ware D, Fulton RS, Stein JC, Wei F, Pasternak S, Liang C, Zhang J, Fulton L, Graves TA, Minx P, Reily AD, Courtney L, Kruchowski SS, Tomlinson C, Strong C, Delehaunty K, Fronick C, Courtney B, Rock SM, Belter E, Du F, Kim K, Abbott RM, Cotton M, Levy A, Marchetto P, Ochoa K, Jackson SM, Gillam B, Chen W, Yan L, Higginbotham J, Cardenas M, Waligorski J, Applebaum E, Phelps L, Falcone J, Kanchi K, Thane T, Scimone A, Thane N, Henke J, Wang T, Ruppert J, Shah N, Rotter K, Hodges J, Ingenthron E, Cordes M, Kohlberg S, Sgro J, Delgado B, Mead K, Chinwalla A, Leonard S, Crouse K, Collura K, Kudrna D, Currie J, He R, Angelova A, Rajasekar S, Mueller T, Lomeli R, Scara G, Ko A, Delaney K, Wissotski M, Lopez G, Campos D, Braidotti M, Ashley E, Golser W, Kim H, Lee S, Lin J, Dujmic Z, Kim W, Talag J, Zuccolo A, Fan C, Sebastian A, Kramer M, Spiegel L, Nascimento L, Zutavern T, Miller B, Ambroise C, Muller S, Spooner W, Narechania A, Ren L, Wei S, Kumari S, Faga B, Levy MJ, McMahan L, Van Buren P, Vaughn MW, Ying K, Yeh C-T, Emrich SJ, Jia Y, Kalyanaraman A, Hsia A-P, Barbazuk WB, Baucom RS, Brutnell TP, Carpita NC, Chaparro C, Chia J-M, Deragon J-M, Estill JC, Fu Y, Jeddeloh JA, Han Y, Lee H, Li P, Lisch DR, Liu S, Liu Z, Nagel DH, McCann MC, SanMiguel P, Myers AM, Nettleton D, Nguyen J, Penning BW, Ponnala L, Schneider KL, Schwartz DC, Sharma A, Soderlund C, Springer NM, Sun Q, Wang H, Waterman M, Westerman R, Wolfgruber TK, Yang L, Yu Y, Zhang L, Zhou S, Zhu Q, Bennetzen JL, Dawe RK, Jiang J, Jiang N, Presting GG, Wessler SR, Aluru S, Martienssen RA, Clifton SW, McCombie WR, Wing RA, Wilson RK (2009) The B73 maize genome: complexity, diversity, and dynamics. Science 326:1112–1115

Tam SM, Causse M, Garchery C, Burck H, Mhiri C, Grandbastien MA (2007) The distribution of copia-type retrotransposons and the evolutionary history of tomato and related wild species. J Evol Biol 20:1056–1072

Wang W, Zheng H, Fan C, Li J, Shi J, Cai Z, Zhang G, Liu D, Zhang J, Vang S, Lu Z, Wong GK, Long M, Wang J (2006a) High rate of chimeric gene origination by retroposition in plant genomes. Plant Cell 18:1791–1802

Wang Y, Tang X, Cheng Z, Mueller L, Giovannoni J, Tanksley SD (2006b) Euchromatin and pericentromeric heterochromatin: comparative composition in the tomato genome. Genetics 172:2529–2540

Wang Y, Diehl A, Wu F, Vrebalov J, Giovannoni J, Siepel A, Tanksley SD (2008) Sequencing and comparative analysis of a conserved syntenic segment in the *Solanaceae*. Genetics 180:391–408

Xiao H, Jiang N, Schaffner EK, Stockinger EJ, van der Knaap E (2008) A retrotransposon-mediated gene duplication underlies morphological variation of tomato fruit. Science 319:1527–1530

Yang L, Bennetzen JL (2009) Distribution, diversity, evolution, and survival of Helitrons in the maize genome. Proc Natl Acad Sci USA 106:19922–19927

Zhong S, Joung J-G, Zheng Y, Chen Y-R, Liu B, Shao Y, Xiang JZ, Fei Z, Giovannoni JJ (2011) High-throughput Illumina strand-specific RNA sequencing library preparation. Cold Spring Harb Protoc 2011(8):940–949, pdb.prot5652