Nonautonomous Long Terminal Repeat Retrotransposons in Plants: Progress and Perspective

Yue Guo[1], Jing Liu[2], Jianchang Du[3*]
Institute of Biotechnology, Provincial Key Laboratory of Agrobiology, Jiangsu Academy of Agricultural Sciences, Nanjing, China

Abstract**—** *Long terminal repeat retrotransposons (LTR-RTs), the major genomic components in plants, can be classified into autonomous and nonautonomous elements based on their internal structures and retrotranspositional properties. Large numbers of nonautonomous elements have been identified, but the factors, and mechanisms that govern their retrotranspositional processes are poorly understood. Here we summarize the recent advance of LTR-RTs in plants, and discuss how nonautonomous LTR-RTs were generated, proliferated and evolved in their host genomes, with an emphasis on the discussion of the partnership and interaction between nonautonomous elements and their autonomous partners. Thus this review will provide insights into the evolution of nonautonomous LTR-RTs, and facilitate our full understanding of the retrotranspositional process of LTR-RTs in plants.*

Keywords**— Retrotansposons, long terminal repeats, evolution, plants.**

I. INTRODUCTION

Retrotransposons are a class of mobile elements, which initiate their retrotransposition through a copy-and-paste mechanism via RNA intermediates. Retrotransposons can be divided into five orders on the basis of their structural features, long terminal repeat retrotransposons (LTR-RTs), *DIRS*-like elements, *Penelope*-like elements (PLEs), LINEs and SINEs [1]. Among them, LTR-RTs are the major genomic components in flowering plants, particularly in species with complex genomes. For example, ~20% of rice [2], ~42% of soybean [3], ~55% of sorghum [4], and >75% of the maize genomes [5] are composed of LTR-RTs. Recent studies indicate that in diploid species, genome size and TE content show a strong positive correlation [6].

A typical LTR-RT element contains two identical LTRs, a primer-binding site (PBS), a polypurine tract (PPT), *gag,* and *pol*, two genes necessary for retrotranspositional process [7]. The LTR region can be further divided into three parts, including U3, R and U5 [7]. Because two LTRs of an element are identical at the time of insertion, the insertion time can be estimated based on the divergence time of the two LTRs and the evolutionary rate of LTR sequences [8]. For example, the majority of LTR-RTs in soybean were amplified in the last 1 million years (Mys) [9]. Usually LTR-RTs are subclassified into *Copia* and *Gypsy* superfamilies based on the order of IN and RT in *pol* [10]. Occasionally some elements were found to contain an additional ORF1 gene upstream of *gag*, and/or envelope (*env*)-like gene after the *pol*. Because both genes are not required for the retrotranspositional process, their origin and functional role remains mysterious. Recent genome-wide analysis and multi-specific comparisons revealed that these elements were anciently evolved, and lineage-specific [1, 9]. Besides intact LTR-RTs, a large number of Solo-LTRs and truncated elements have also been found in plant genomes [9, 11, 12]. These incomplete elements, together with numerous LTR remnants were presumed to be the products of unequal recombination and illegitimate recombination, two molecular mechanisms counterbalancing genome expansion [11, 12]. For instance, it was estimated that >190 Mb of DNA had been removed from the rice genome in the past 8 Mys, leaving the current rice genome ~400 Mb with ~97 Mb DNA of detectable LTR-RTs [12].

Based on their structural completeness and retrotranspositional capability, LTR-RTs can also be classified into autonomous and nonautonomous types. An intact element is defined as autonomous if it encodes all the protein-coding domains necessary for catalyzing its retrotransposition [1]. By contrast, an element lacking one or more protein coding domains, but still keeping its retrotranspositional activity within a time frame, is generally defined as nonautonmous. Large retrotransposon derivatives (LARDs) and terminal-repeat retrotransposons in miniature (TRIM) are two groups of LTR-RTs belonging to nonautonomous types [13, 14]. Since both LARDs and TRIM have no open reading frames in the internal part, they were presumed to transpose by borrowing proteins from their autonomous partners. But for most cases, the relationships between autonomous and nonautonomous elements have not been established yet, and the exact mechanism(s) governing the activity of nonautonomous elements remains unclear. Although the transpositional mechanism of *Ds1* in maize, and MITEs in rice have been discussed previously [15-19], these nonautonomous elements transpose via a "cut-and-paste" mechanism, and do not undergo reverse transcription process. Thus the transposition mechanism for these elements may be essentially different from those of LTR-RTs.

We have previously identified 510 LTR-RT families in the sequenced soybean genome, and conducted further comprehensive analysis of the largest family *SNARE* [20, 21]. This family contains both autonomous and nonautonomous subfamilies. We found that nonautonomous elements frequently exchanged the LTR domains with their autonomous partners in different timeframes of soybean evolutionary history, thus providing the evidence that autonomous and nonautonomous LTR-RTs can interact and communicate with each other. Here we review the recent studies on plant nonautonomous elements with respect to the nature, timing, origin, and evolutionary process, thereby providing insights into the retrotranspositional process of LTR-RTs.

II. STRUCTURAL FEATURES OF NONAUTONOMOUS LTR-RTS

Nonautonomous LTR-RTs are widespread throughout eukaryotic lineages, particularly in flowering plants [22-24]. There are two types of nonautonomous elements based on the absence or presence of coding genes in the middle. One type includes both LARDs and TRIM, neither of which contains any signature of retrotranspositional related genes, such as *gag*, and *pol* [13, 14]. In LARD elements, the coding region is replaced by a long, conserved noncoding domain. Nevertheless, the internal parts of TRIM is almost completely lacking, and the LTRs are quite short, making the intact element very small (<1 kb, [13]). The second type includes some nonautonomous elements, such as *nonCRR1*/*nonCRR2* [25], *Retand-1*[26], *BARE-2* [27, 28], and *SNRE* [21]. The elements in this type have detectable *gag* and/or *pol*, but the coding regions have been either highly degraded, or disrupted by frameshifts and stop codons, indicating that these elements are defective, and require products from other elements in trans to amplify in their host genome. The majority, if not all, of the nonautonomous elements investigated thus far, have two intact LTR sequences, where initiation and termination sites resides in, PBS, and PPT. It may reflect the minimum information for nonautonomous elements to move in a genome. It was frequently observed that tandem repeats (usually 24~100 bp for each monomer) exist in the internal parts, particularly in the regions between the *pol* and 3' LTR [21, 26, 29]. In spite of the conserved location of tandem repeats, little is known regarding their origin and functional role.

III. THE ABUNDANCE, TIMING AND ORIGIN OF NONAUTONOMOUS ELEMENTS

Although nonautonomous elements have been frequently identified, their abundance, timing, nature, and origin are not well understood. Previous studies revealed that some nonautonomous families, such as *Dasheng* elements in rice, have a few hundred copies, most of which have identical LTRs, indicating that they were amplified quite recently [29]. In contrast, the maize family *Zeon-1* is one of the oldest families in their host genome [30]. Interestingly, although the nonautonomous LTR subfamily *SNRE* in soybean have bursts at ~2 Mys, a subset of these elements with a piggy-backing Solo-LTR were dramatically amplified within the last 0.5 Mys [9]. Taken together, these observations indicate that the timing and amplification of nonautonomous elements are variable across species and families, and are affected by their host genomes, different evolutionary history, and nature selection on the genes involved in retrotranspositional process [31].

The origin of nonautonomous elements remains mysterious. Sequence comparisons of centromere RTs revealed that *noaCRR* and *CRR* elements in rice share substantial sequence similarity of the LTRs, and conserved motifs, including two terminal ends of LTR sequences, PBS, and PPT sites, indicating that *noaCRR* elements were likely derived from *CRR* [25]. For soybean *SNARE* family, nonautonomous subfamily *SNRE* elements and autonomous subfamily *SARE*^{A} elements share the same type of tandem repeats, and the majority of *SNRE* elements and *SARE*^{A} elements were phylogenetically clustered in a monophyletic group [21]. Thus it is reasonable to deduce that *SNRE* elements were derived from *SARE*^{A} instead of *SARE*^{B} [21].

IV. RETROTRANSPOSITION AND AMPLIFICATION PROCESS OF NONAUTONOMOUS ELEMENTS

Because nonautonomous elements do not have a full set of genes encoding proteins necessarily for retrotransposition, these elements are supposed to use the same or very similar enzyme machinery with their autonomous partners. Using a homology-based approach, the putative autonomous and nonautonomous partners, *RIRE2* and *Dasheng* were identified in the rice genome [29]. But basically *RIRE2* and *Dasheng* elements were grouped into two distinct clades based on the LTR sequences, suggesting few recombination events occurred between *RIRE2* and *Dasheng* families. In the rice genome, four other possible partnerships had been established, such as *nonCRR1*/*CRR1*, *nonCRR2*/*CRR2*, *spip*/*RIRE3*, and *Squiq*/*RIRE8* (Table 1, [25, 32]). For example, in the *S. latifolia* genome, *Rend-1* (nonautonomous) elements were found similarly abundant as *Rend-2* elements (autonomous), and were widely transcribed in all tissues tested [26]. Like *RIRE2* and *Dasheng*, the above partners belonged to *Gypsy*-like superfamily, and were basically separated into two distinct families. However, a *Copia*-like family *BARE-2* and its putative autonomous partners *BARE-1*/*Wis-2* were identified in *H. vulgare* (Table 1, [27]). More interestingly, *BARE-2* appears to be a chimeric element, because the two LTRs and 5' UTR regions of *BARE-2* were more similar to *BARE-1*, whereas the rest are more similar to another family *Wis-2*. The chimeric structure was likely generated from retrotransposon recombination by strand switching during replication [27].

TABLE 1
SUMMARY OF THE NONAUTONOMOUS FAMILIES AND THEIR PUTATIVE AUTONOMOUS PARTNERS

Species	Superfamily	Nonautonomous	Predicted autonomous	Reference
Oryza sativa	*Gypsy*-like	*noaCRR1*	*CRR1*	[25]
Oryza sativa	*Gypsy*-like	*noaCRR2*	*CRR2*	[25]
Oryza sativa	*Gypsy*-like	*Dasheng*	*RIRE2*	[29]
Oryza sativa	*Gypsy*-like	*Spip*	*RIRE3*	[32]
Oryza sativa	*Gypsy*-like	*Squiq*	*RIRE8*	[32]
Silene latifolia	*Gypsy*-like	*Retand-1*	*Retand-2*	[26]
Hordeum vulgare	*Copia*-like	*BARE-2*	*BARE-1/Wis-2*	[27]
Glycine max	*Gypsy*-like	*SNRE*	*SARE*	[21]

More data regarding RNA level recombination comes from a recent study of *SNARE* family in soybean (Table 1, [21]). *SNARE* family contains two autonomous subfamilies *SARE*^{A} and *SARE*^{B}, and a nonautonomous subfamily *SNRE*. Unexpectedly a subset of *SNRE* elements (called as *SNRE*^{S}) carry a foreign Solo-LTR at the same site in the internal region. Nonautonomous subfamily *SNRE* elements and autonomous subfamily *SARE* share highly identical LTR sequences, identical PBS and PPT sites, conserved tandem repeats, similar distribution pattern, and preferential integration sites (overall bias for G or C). Furthermore, phylogenetic analysis and case-based structural examination between the recombinants and their parental elements revealed that, extensive region-specific sequences have swapped within the recent evolutionary timeframes. The majority of the recombinants were difficult to be explained by genomic recombination. In contrast, the recombinant LTR structures were more consistent with the RNA recombination model. If the genomic recombination model holds true, the new copies will contain chimeric LTR sequences after amplification. But the data showed that the whole LTR regions shared highly similarity with that of *SARE*, and the coding regions were more similar to *SNRE* [21]. In summary, all these observations indicate that, RNA level recombination rather than genomic recombination mediates *SNARE* evolution, and that a molecular mechanism may be involved in the enhancement between autonomous and nonautonomous elements [21].

V. CONCLUSION

With more genomic sequences are available, many nonautonomous retrotransposons will be identified and annotated. Because almost all eukaryote genomes contains transposable elements, further understanding of the structure, evolution and replication of nonautonomous elements will be helpful to decipher their important roles in the host genome evolution. Perhaps more work will be focused on how and what frequency the nonautonomous elements interplay and communicate with their partners, and how they amplify in the genome, regulate gene expression, and drive the host genome evolution.

ACKNOWLEDGEMENTS

This work is supported by the National Natural Science Foundation of China (31370266), Jiangsu Agricultural Science and Technology Innovation Funds (CX(14)5010), the Open Fund from the Key Laboratory of Biology and Genetic Improvement of Oil Crops, Ministry of Agriculture (2016002), and the State Key Laboratory of Cotton Biology Open Fund (CB2016B03).

REFERENCES

[1] T. Wicker, F. Sabot, A. Hua-Van, J. L. Bennetzen, P. Capy, B. Chalhoub, A. Flavell, P. Leroy, M. Morgante, O. Panaud, E. Paux, P. SanMiguel and A. H. Schulman, “A unified classification system for eukaryotic transposable elements,” in Nat. Rev. Genet., 2007, 8, pp 973-982.

[2] Z. Tian, C. Rizzon, J. Du, L. Zhu, J. L. Bennetzen, S. A. Jackson, B. S. Gaut and J. Ma, “Do genetic recombination and gene density shape the pattern of DNA elimination in rice long terminal repeat retrotransposons?” in Genome Res., 2009, 19, pp 2221-2230.

[3] J. Schmutz, S. B. Cannon, J. Schlueter, J. Ma, T. Mitros, W. Nelson, D. L. Hyten, Q. Song, J. J. Thelen, J. Cheng, D. Xu, U. Hellsten, G. D. May, Y. Yu, T. Sakurai, T. Umezawa, M. K. Bhattacharyya, D. Sandhu, B. Valliyodan, E. Lindquist, M. Peto, D. Grant, S. Shu, D. Goodstein, K. Barry, M. Futrell-Griggs, B. Abernathy, J. Du, Z. Tian, L. Zhu, N. Gill, T. Joshi, M. Libault, A. Sethuraman, X. C. Zhang, K. Shinozaki, H. T. Nguyen, R. A. Wing, P. Cregan, J. Specht, J. Grimwood, D. Rokhsar, G. Stacey, R. C. Shoemaker and S. A. Jackson, “Genome sequence of the palaeopolyploid soybean,” in Nature, 2010, 463, pp 178-183.

[4] A. H. Paterson, J. E. Bowers, R. Bruggmann, I. Dubchak, J. Grimwood, H. Gundlach, G. Haberer, U. Hellsten, T. Mitros, A. Poliakov, J. Schmutz, M. Spannagl, H. Tang, X. Wang, T. Wicker, A. K. Bharti, J. Chapman, F. A. Feltus, U. Gowik, I. V. Grigoriev, E. Lyons, C. A. Maher, M. Martis, A. Narechania, R. P. Otillar, B. W. Penning, A. A. Salamov, Y. Wang, L. Zhang, N. C. Carpita, M. Freeling, A. R. Gingle, C. T. Hash, B. Keller, P. Klein, S. Kresovich, M. C. McCann, R. Ming, D. G. Peterson, R. Mehboob ur, D. Ware, P. Westhoff, K. F. Mayer, J. Messing and D. S. Rokhsar, “The Sorghum bicolor genome and the diversification of grasses,” in Nature, 2009, 457, pp 551-556.

[5] P. S. Schnable, et al., “The B73 maize genome: complexity, diversity, and dynamics,” in Science, 2009, 326, pp 1112-1115.

[6] M. I. Tenaillon, J. D. Hollister and B. S. Gaut, “A triptych of the evolution of plant transposable elements,” in Trends Plant Sci., 2010, 15, pp 471-478.

[7] A. Kumar and J. L. Bennetzen, “Plant retrotransposons,” in Annu. Rev. Genet., 1999, 33, pp 479-532.

[8] P. SanMiguel, B. S. Gaut, A. Tikhonov, Y. Nakajima and J. L. Bennetzen, “The paleontology of intergene retrotransposons of maize,” in Nat. Genet., 1998, 20, pp 43-45.

[9] J. Du, Z. Tian, C. S. Hans, H. M. Laten, S. B. Cannon, S. A. Jackson, R. C. Shoemaker and J. Ma, “Evolutionary conservation, diversity and specificity of LTR-retrotransposons in flowering plants: insights from genome-wide analysis and multi-specific comparison,” in Plant J., 2010, 63, pp 584-598.

[10] Y. Xiong and T. H. Eickbush, “Origin and evolution of retroelements based upon their reverse transcriptase sequences,” in Embo J., 1990, 9, pp 3353-3362.

[11] K. M. Devos, J. K. Brown and J. L. Bennetzen, “Genome size reduction through illegitimate recombination counteracts genome expansion in Arabidopsis,” in Genome Res., 2002, 12, pp 1075-1079.

[12] J. Ma, K. M. Devos and J. L. Bennetzen, “Analyses of LTR-retrotransposon structures reveal recent and rapid genomic DNA loss in rice,” in Genome Res., 2004, 14, pp 860-869.

[13] C. P. Witte, Q. H. Le, T. Bureau and A. Kumar, “Terminal-repeat retrotransposons in miniature (TRIM) are involved in restructuring plant genomes,” in Proc. Natl. Acad. Sci. USA, 2001, 98, pp 13778-13783.

[14] R. Kalendar, C. M. Vicient, O. Peleg, K. Anamthawat-Jonsson, A. Bolshoy and A. H. Schulman, “Large retrotransposon derivatives: abundant, conserved but nonautonomous retroelements of barley and related genomes,” in Genetics, 2004, 166, pp 1437-1450.

[15] S. R. Wessler, “The maize transposable Ds1 element is alternatively spliced from exon sequences,” in Mol. Cell Biol., 1991, 11, pp 6192-6196.

[16] N. Jiang, Z. Bao, X. Zhang, H. Hirochika, S. R. Eddy, S. R. McCouch and S. R. Wessler, “An active DNA transposon family in rice,” in Nature, 2003, 421, pp 163-167.

[17] N. Jiang, C. Feschotte, X. Zhang and S. R. Wessler, “Using rice to understand the origin and amplification of miniature inverted repeat transposable elements (MITEs),” in Curr. Opin. Plant Biol., 2004, 7, pp 115-119.

[18] X. Zhang, N. Jiang, C. Feschotte and S. R. Wessler, “PIF- and Pong-like transposable elements: distribution, evolution and relationship with Tourist-like miniature inverted-repeat transposable elements,” in Genetics, 2004, 166, pp 971-986.

[19] G. Yang, D. H. Nagel, C. Feschotte, C. N. Hancock and S. R. Wessler, “Tuned for transposition: molecular determinants underlying the hyperactivity of a Stowaway MITE,” in Science, 2009, 325, pp 1391-1394.

[20] J. Du, D. Grant, Z. Tian, R. T. Nelson, L. Zhu, R. C. Shoemaker and J. Ma, “SoyTEdb: a comprehensive database of transposable elements in the soybean genome,” in BMC Genomics, 2010, 11, 113.

[21] J. Du, Z. Tian, N. J. Bowen, J. Schmutz, R. C. Shoemaker and J. Ma, “Bifurcation and enhancement of autonomous-nonautonomous retrotransposon partnership through LTR Swapping in soybean,” in Plant Cell, 2010, 22, pp 48-61.

[22] E. M. McCarthy, J. Liu, G. Lizhi and J. F. McDonald, “Long terminal repeat retrotransposons of Oryza sativa,” in Genome Biol., 2002, 3, RESEARCH0053.

[23] F. Sabot and A. H. Schulman, “Parasitism and the retrotransposon life cycle in plants: a hitchhiker's guide to the genome,” in Heredity, 2006, 97, pp 381-388.

[24] A. Wawrzynski, T. Ashfield, N. W. Chen, J. Mammadov, A. Nguyen, R. Podicheti, S. B. Cannon, V. Thareau, C. Ameline-Torregrosa, E. Cannon, B. Chacko, A. Couloux, A. Dalwani, R. Denny, S. Deshpande, A. N. Egan, N. Glover, S. Howell, D. Ilut, H. Lai, S. M. Del Campo, M. Metcalf, M. O'Bleness, B. E. Pfeil, M. B. Ratnaparkhe, S. Samain, I. Sanders, B. Segurens, M. Sevignac, S.

Sherman-Broyles, D. M. Tucker, J. Yi, J. J. Doyle, V. Geffroy, B. A. Roe, M. A. Maroof, N. D. Young and R. W. Innes, “Replication of nonautonomous retroelements in soybean appears to be both recent and common,” in Plant Physiol., 2008, 148, pp 1760-1771.

[25] K. Nagaki, P. Neumann, D. Zhang, S. Ouyang, C. R. Buell, Z. Cheng and J. Jiang, “Structure, divergence, and distribution of the CRR centromeric retrotransposon family in rice,” in Mol. Biol. Evol., 2005, 22, pp 845-855.

[26] E. Kejnovsky, Z. Kubat, J. Macas, R. Hobza, J. Mracek and B. Vyskot, “Retand: a novel family of gypsy-like retrotransposons harboring an amplified tandem repeat,” in Mol. Genet. Genomics, 2006, 276, pp 254-263.

[27] C. M. Vicient, R. Kalendar and A. H. Schulman, “Variability, recombination, and mosaic evolution of the barley BARE-1 retrotransposon,” J. Mol. Evol., 2005, 61, pp 275-291.

[28] J. A. Tanskanen, F. Sabot, C. Vicient and A. H. Schulman, “Life without GAG: the BARE-2 retrotransposon as a parasite's parasite,” in Gene, 2007, 390, pp 166-174.

[29] N. Jiang, I. K. Jordan and S. R. Wessler, “Dasheng and RIRE2. A nonautonomous long terminal repeat element and its putative autonomous partner in the rice genome,” in Plant Physiol., 2002, 130, pp 1697-1705.

[30] W. Hu, O. P. Das and J. Messing, “Zeon-1, a member of a new maize retrotransposon family,” in Mol. Gen. Genet., 1995, 248, pp 471-480.

[31] R. S. Baucom, J. C. Estill, J. Leebens-Mack and J. L. Bennetzen, “Natural selection on gene function drives the evolution of LTR retrotransposon families in the rice genome,” in Genome Res., 2009, 19, pp 243-254.

[32] C. Vitte, C. Chaparro, H. Quesneville and O. Panaud, “Spip and Squiq, two novel rice non-autonomous LTR retro-element families related to RIRE3 and RIRE8,” in Plant Sci., 2007, 172, pp 8-19.