RESEARCH

SIGNPOST

Retrovirus-like elements in plants

Antonio Marco and Ignacio Marín
Departamento de Genética, Universidad de Valencia, Calle Doctor Moliner, 50
46100 Burjassot (Valencia), Spain

Abstract

LTR retrotransposons with structures that are identical to those found in simple vertebrate retroviruses, including a putative env gene, have been discovered in plants. Those potential plant retroviruses can be classified into two classes. The first one is formed by the Arabidopsis thaliana Athila elements and many other closely related env-containing elements. All of them belong to the Ty3/Gypsy group of LTR retrotransposons. The second class, in which the best-known element was first found in soybean and called SIRE1, belongs to the Ty1/Copia group. Thus, two distantly related lineages have convergent features that suggest that the transition between intracellular and infective ways of life may have occurred several times independently.

Correspondence/Reprint request: Dr. Ignacio Marín, Departamento de Genética, Universidad de Valencia
Calle Doctor Moliner, 50, 46100 Burjassot (Valencia), Spain. E-mail: ignacio.marin@uv.es

Introduction

First discovered in maize by Barbara McClintock and later recognized in all types of organisms, transposable elements are ubiquitous components of plant genomes. Eukaryotic transposable elements are generally divided into two classes. Class II elements or DNA transposons do not require an intermediate RNA. Class I elements however, require an intermediate RNA and cannot complete its replicative cycle without the use of the enzyme reverse transcriptase (RT), able to generate DNA molecules from RNA templates. Both Class I and Class II elements are present in plant species [1,2]. In fact, a particular type of Class I elements, the retrotransposons, account for a great amount of genomic DNA in plants, from 15% in *Arabidopsis thaliana* to perhaps 70 – 80% in some species such as maize or barley. In the case of maize genome, the expansion in the past 5 - 6 million years of LTR retrotransposons doubled its genome size (reviewed in [2]). Retrotransposons significantly contribute to the shaping of plant genomes, inducing gene mutations and favouring the production of genomic rearrangements by recombination [1]. However only a small fraction of plant retrotransposons is active in a given genome [2]. In some cases, it has been shown that intra-element recombination among the LTRs counterbalances the accumulation on LTR-retrotransposons, leaving solo-LTRs in the genome (e.g. [3]). This review will be focused on a particular type of Class I elements, the LTR retrotransposons, which are characterized by having a structure very similar or even identical to that of the simplest vertebrate retroviruses (Figure 1).

The name "LTR retrotransposons" refers to their characteristic Long Terminal Repeats (LTRs), that delimit the beginning and the end of the element (see Figure 1). Autonomous LTR retrotransposons, able to replicate independently, contain also at least two genes, called *gag* and *pol*, that encode either for a capside-like protein (Gag) or for a polyprotein that gave rise to the protease, RT, RNAse H (usually as a domain associated with the RT protein) and integrase (IN) activities, required for the completion of the retrotransposon replicative cycle. The main difference between retroviruses and most LTR retrotransposons is that the latter lack a third gene (called *env*) that is found in all retroviruses and encodes for proteins involved in interacting with cellular receptors and mediate fusion of the host and viral membranes [4].

Studies by Thomas Eickbush and coworkers demonstrated that retroviruses and LTR retrotransposons are evolutionarily very close relatives. Xiong and Eickbush [5] found that retrovirus RTs are very similar to those of a particular type of LTR retrotranposons, classically called "Ty3/Gypsy" and recently renamed as "Metaviridae" (Figure 1; [6,7]), while a second major LTR retrotransposon group (the "Ty1/Copia group" or "Pseudoviridae"; structural details also in Figure 1) was more distantly related. These results strongly supported the hypothesis, first formulated by Howard Temin, that
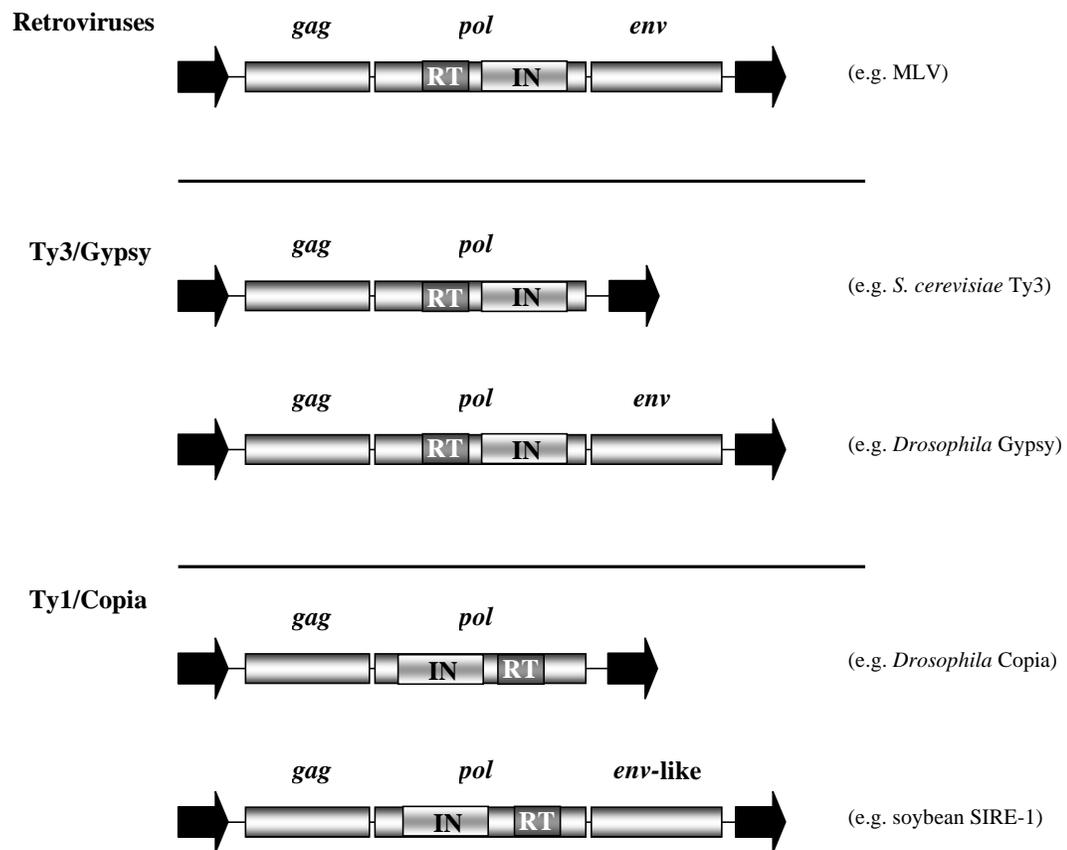
Figure 1. Structures of retroviruses and the two main classes of LTR retrotransposons (Ty3/Gypsy and Ty1/Copia). Arrows correspond to the Long Terminal Repeats. RT: reverse transcriptase. IN: integrase.

retroviruses may have originated from LTR retrotransposons that managed to escape cellular control [8].

Plant LTR retrotransposons with *ENV*-like sequences

The distinction between retrovirus and LTR retrotransposons started to blur when it was noticed that several insect Ty3/Gypsy elements actually contained a third ORF (ORF3) that might encode for an ENV-like protein. Finally, the discovery that one of the best known LTR retrotransposons, the *Drosophila melanogaster* element Gypsy, was able in some circumstances to propagate among cells as an infective retroviruses [9,10] led to the idea that retroviruses and LTR retrotransposons may well be considered the same type of entities, and that occasional transitions from a purely intracellular to an infective mode or life (or viceversa) may occasionally happen. These results suggested the possibility of finding retroviruses in many other organisms. The elements other than retroviruses found to contain an ORF3, and thus potentially infected, were defined as forming a new genus, and called "errantiviruses".

In plants, LTR retrotransposons with an ORF3 were first discovered by Wright and Voytas [11] that described that some copies of an *Arabidopsis thaliana* element called Athila possessed not only classical *gag* and *pol* sequences, but also an additional open reading frame that might encode for an *env*-like protein. Defective Athila elements had been characterized some time before [12], but it was Wright and Voytas' work the one that demonstrated that Athilas are typical Ty3/Gypsy retrotransposons. They discovered Athila RT-containing elements and the phylogenetic analyses of those RT sequences showed their close relationships to other plant Ty3/Gypsy group elements. It became clear also that the insect "errantiviruses" and plant elements with ORF3 were distant relatives [11]. These results already suggested that any "errantivirus" definition based on the presence of an ORF3 was meaningless from an evolutionary point of view, not corresponding to any monophyletic lineage. In spite of the lesson taught by these results, the current official classification of LTR retrotransposons is still plagued by this type of inconsistencies (see summary in [13]).

In the most comprehensive survey performed to that date, we characterized the main Ty3/Gypsy lineages present in *Arabidopsis thaliana*, demonstrating the existence of six families never hitherto defined [14]. One of them was called Little Athila, because of its similarity to Athila elements. Athila and Little Athila elements are also present in *Brassica oleracea*, as recently shown by Zhang and Whesler [15]. The name "Little Athila" derives from the fact that elements of this type with an ORF3 were never observed. Our study also shown that Athilas with and without ORF3 were mixed in the RT-based phylogenetic tree, suggesting that occasional losses and acquisitions of those sequences may have been occurring in the Athila lineages. We also showed that the Athila and Little Athila elements were close relatives to a few known elements found in other dicot species, called Cyclops (isolated from pea and faba bean) and Diaspora (from soybean). The Cyclops elements contained also an ORF3 [16]. Our work was shortly followed by a most interesting study [17] that confirmed our analyses and established that retrotransposons closely related to Athila with *env*-like sequences were present in many other plant species, including both monocots and dicots. Related results were reported by Wright and Voytas [18]. These results suggest that the origin of the *env*-like sequences in Athila elements was previous to the monocot/dicot split, about 200 millions of years ago. Therefore, we can exclude the possibility of these elements being any kind of rare retrovirus that has been recently horizontally transferred to plant genomes from an animal host. Figure 2 shows a recent analysis in which we searched for RT sequences that correspond to Athila-like elements in multiple plant species. The tree is rooted with plant chromoviruses, that are distantly related Ty3/Gypsy elements [14,19]. Here, we report for the first time the presence of Athila-like elements in gymnosperms (*Pinus*),
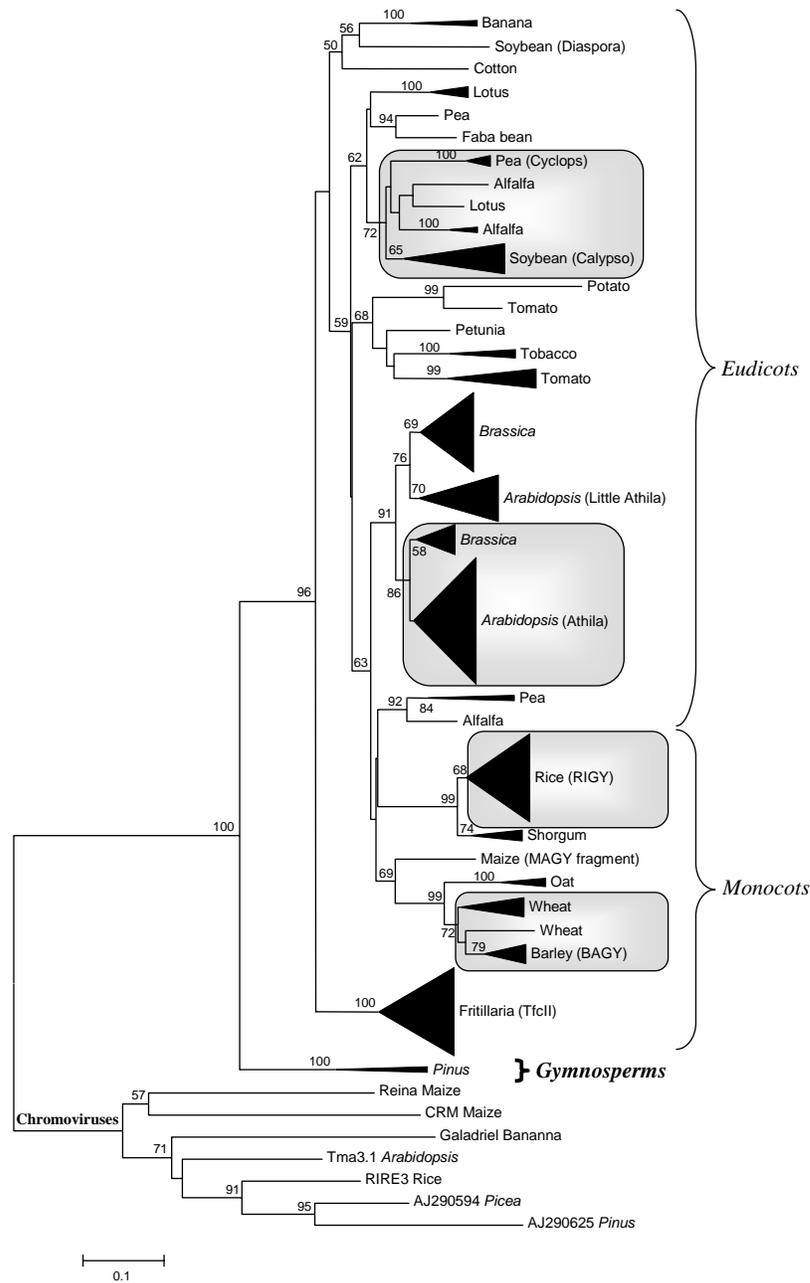
Figure 2. Neighbor-joining tree of Athila-like retrotransposon sequences. Triangles are proportional to the number of elements included. Grey boxes detail characterized elements with *env*. Numbers refer to percentage of support for the branches, using bootstrap analysis.

establishing that these elements have a very ancient origin in plant species. Whether an exhaustive search in gymnosperms may unearth Athila elements with an ORF3 is an interesting, still unsolved, question.

In parallel to these discoveries, a significant second line of research was initiated by the discovery that a soybean (*Glycine max*) element, SIRE1, also

contains putative *env* sequences [20]. The main importance of this finding regards the fact that SIRE1 is not a Ty3/Gypsy element, but it belongs instead to the distant Ty1/Copia group. This result strongly suggested that totally unrelated plant LTR retrotransposons may have independently acquired *env* sequences, potentially being converted into two different types of plant retroviruses (see review in [21]). A recent report demonstrates that SIRE1 active elements are abundant in soybean, and that all elements of this family are closely related, suggesting a relatively recent amplification [22].

An obvious question concerns the origin of the *env*-like sequences found in those plant retrotransposons. It is reasonable to think that they may have been co-opted, captured from some foreign source. In this direction, the work by Malik and coworkers [23] that suggests that insect and nematode retrotransposons may have *env* sequences derived from proteins found in different types of viruses is most significant. A similar origin for vertebrate retrovirus *env*s, although perhaps from combination of modules derived from more than a single source, has been also postulated (reviewed in [24]). However, so far, the particular source that provided the Athila or SIRE1 *env*-like proteins has not been determined.

Are they real *ENV* genes?

A major problem in this context is to establish whether the *env*-like sequences actually encode for ENV proteins or they encode instead for some other type of proteins, totally unrelated with retroviral ENV function. A second, related question is whether those ENV proteins are actually used to generate infective plant retroviruses. These questions will remain, in our opinion, unsolved until a functional test may be devised to determine the function of the putative ENVs *in vivo*. So far, the only available evidence are the analyses by the authors that have characterized the *env*-like sequences, devised to demonstrate that the biochemical features of the deduced proteins are compatible with what is known of the function of ENVs in retroviruses [11,18,20]. In this sense, much emphasis has been placed on these putative proteins containing, as happens in retroviral ENVs, transmembrane domains. However, those similarities may be significant, demonstrating a common mode of action of both types of proteins, or just be a convergent feature, with retroviral ENVs and plant ENV-like proteins having totally unrelated roles. In fact, the plant cell wall is an important obstacle for cell to cell infection by enveloped viruses, making unclear whether retrovirus may at all exist in plants. A first suggestion to circumvent this problem was that plant retroviruses might disseminate through plasmodesmata [25]. If this is indeed the case, it would be reasonable to predict that the ENV-like proteins would have very different roles from retroviral ENVs. A second suggestion was that plant retroviruses may have life cycles similar to those of some other *env*-containing plant

viruses, that shift between animal and plant hosts [20]. These possibilities remain speculative.

Evolutionary success of ENV-containing elements

We have already discussed that Athila-related *env*-containing elements can be found in many species. This success in spreading and colonizing multiple genomes has a parallel reflection in the fact that many of those elements are also very abundant in the genomes that they colonize. Vicient et al. [17] summarized the information available for some of these elements, estimating that the number of copies for Athila-related elements ranges from hundreds in *Arabidopsis thaliana* to tens of thousands in rice. SIRE1 is also found in about 1000 copies in soybean [22]. In any case, to determine with precision the total number of elements is difficult, because there are many truncated or highly diverged copies. Interestingly, a second problem may be the intrinsic heterogeneity of the structures of the elements. For example, we have recently found that to search for Athila elements using RTs sequences is dramatically misleading. In *Arabidopsis*, most Athila elements lack *pol* sequences, containing only LTRs plus *gag*s or, alternatively, LTRs, *gag* and *env* sequences (Figure 3). Figure 4 shows the result of a phylogenetic analysis based on the Gag proteins, the only coding sequences common to most Athilas. As it is shown in those figures, only a few Athila lineages have RTs. We are currently examining the reasons for this peculiar distribution and the evolutionary dynamics that it may generate.
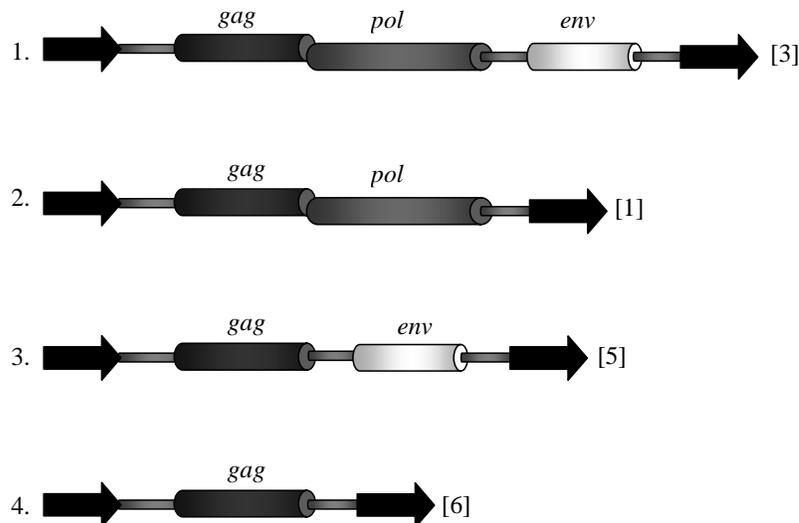
Figure 3. The four structural types of Athila retrotransposons found in the *Arabidopsis thaliana* genome. Notice that all elements have *gag* sequences, but many lack *pol* sequences, and thus cannot independently replicate. Number in brackets refers to how many clades have these structures (defined as in Figure 4).

Figure 4. Topology of the phylogenetic tree for the Athila retrotransposons in *Arabidopsis thaliana* based on *gag* (ORF1) amino acidic sequences. Numbers above the branches refer to bootstrap support. Lineages in gray contain *pol* sequences, while most lineages, shown in black, do not have them. Numbers in bold indicate the structure of the elements, according to the four classes detailed in Figure 3.

Vicient et al. [26] demonstrated that ESTs of Athila-like elements can be found, suggesting that they are transcriptionally active, although at low levels, in multiple species. The low level of transcription is most probably associated with the elements being under transcriptional gene silencing (TGS). It has been shown that mutations in genes involved in TGS dramatically increase Athila

retrotransposons expression ([27]; reviewed in [2]). In any case, it is evident that both Athila-like and SIRE1 elements are currently active in most/all genomes in which they are found.

The future of plant retroviruses research

If indeed it is finally shown that the elements that we have detailed in this summary behave as infective retroviruses, this may open many new possibilities of genetic manipulation in plants. Currently, retroviral vectors are widely used to transfer genes in vertebrate animals and similar systems could be devised to be used in plant species. On the other hand, retroviruses might contribute to explain cases of horizontal transfer to and from plant genomes, a question that deserves independent consideration. We think that the completion of sequencing projects of plant genomes may provide important clues required for deciphering the origin and evolution of retrotransposons and retroviruses.

References

1. Kumar, A., and Bennetzen, J.F. 1999, Annu. Rev. Genet., 33, 479.
2. Feschotte, C., Jiang, N. and Wessler, S.R. 2002, Nat. Rev. Genet., 3, 329.
3. Vitte, C., and Panaud, O. 2003, Mol. Biol. Evol., 20, 528.
4. Swanstrom, R., and Wills, J.W. 1997, Synthesis, assembly, and processing of viral proteins, J.M. Coffin et al (Eds.), Cold Spring Harbor Laboratory Press. New York.
5. Xiong, Y., and Eickbush, T.H. 1990, EMBO J., 9, 3353.
6. Eickbush, T.H. 1994, Origin and evolutionary relationships of retroelements, S. S. Morse (Ed.), Raven Press Ltd., New York. 121.
7. Eickbush T.H., and Malik H.S. 2002, Origins and evolution of retrotransposons, N.L. Craig et al (Ed.), ASM Press, Washington, 1111.
8. Temin, H.M. 1980, Cell, 21,599.
9. Kim, F.J., Battini, J L., Manel, N., and Sitbon, M. 2004, Virology, 318, 183.
10. Song, S.U., Gerasimova, T., Kurkulos, M., Boeke, J.D., and Corces, V.G. 1994, Genes Dev., 8, 2046.
11. Wright, D.A., and Voytas, D.F. 1998, Genetics, 149, 703.
12. Pélissier, T., Tutois, S., Deragon, J.M., Tourmente, S., Genestier, S., and Picard, G. 1995, Plant Mol. Biol., 29, 441.
13. Havecker, E.R., Gao, X., and Voytas, D.F. 2004, Genome Biol., 5, 225.
14. Marín, I. and Lloréns, C. 2000, Mol. Biol. Evol., 17, 1040.
15. Zhang, X. and Wessler, S.R. 2004, Proc. Natl. Acad. Sci. USA, 101, 5589.
16. Chavanne, F., Zhang, D.X., Liaud, M.F., and Cerff, R. 1998, Plant Mol. Biol., 37, 363.
17. Vicient, C.M., Kalendar, R. and Schulman, A.H. 2001, Genome Res., 11, 2041-2049.
18. Wright, D.A., and Voytas, D.F. 2002, Genome Res., 12, 122.
19. Gorinsek, B., Gubensek, F., and Kordis, D. 2004, Mol. Biol. Evol., 21, 781.
20. Laten, H.M., Majumdar, A., and Gaucher, E.A. 1998, Proc. Natl. Acad. Sci. USA, 95, 6897.

21. Peterson-Burch, B.D., Wright, D.A., Laten, H.M. and Voytas, D.F. 2000, Trends Genet., 16, 151.
22. Laten, H.M., Havecke, E.R., Farmer, L.M., and Voytas, D.F. 2003, Mol. Biol. Evol., 20, 1222.
23. Malik, H.S., Henikoff, S., and Eickbush, T.H. 2000, Genome Res., 10, 1307.
24. Kim, A. et al. 1994, Proc. Natl. Acad. Sci. USA, 91, 1285.
25. Kumar, A. 1998, Trends Plant Sci., 3,371.
26. Vicient, C.M., Jääskeläinen, Kalendar, R. and Schulman, A.H. 2001, Plant Physiol., 125,1283.
27. Steimer, A., Amedeo, P., Afsar, K., Fransz, P., Scheid, O.M., and Paszkowski, J. 2000, Plant Cell, 12,1165.