

Retroviruses in plants?

Eukaryotic genomes harbor mobile genetic elements known as long terminal repeat (LTR) retrotransposons. LTR retrotransposons are closely related to the infectious and endogenous retroviruses, and they are collectively referred to as LTR retroelements. LTR retrotransposons and retroviruses have two genes in common – *gag*, which encodes proteins that assemble into virus or virus-like particles, and *pol*, which encodes the enzymatic activities required for replication. The *envelope* (*env*) gene of the retroviruses distinguishes them from the LTR retrotransposons. Envelope glycoproteins associate with cell membranes and facilitate the budding of viral core particles from infected cells. They also mediate infection by recognizing cellular receptors.

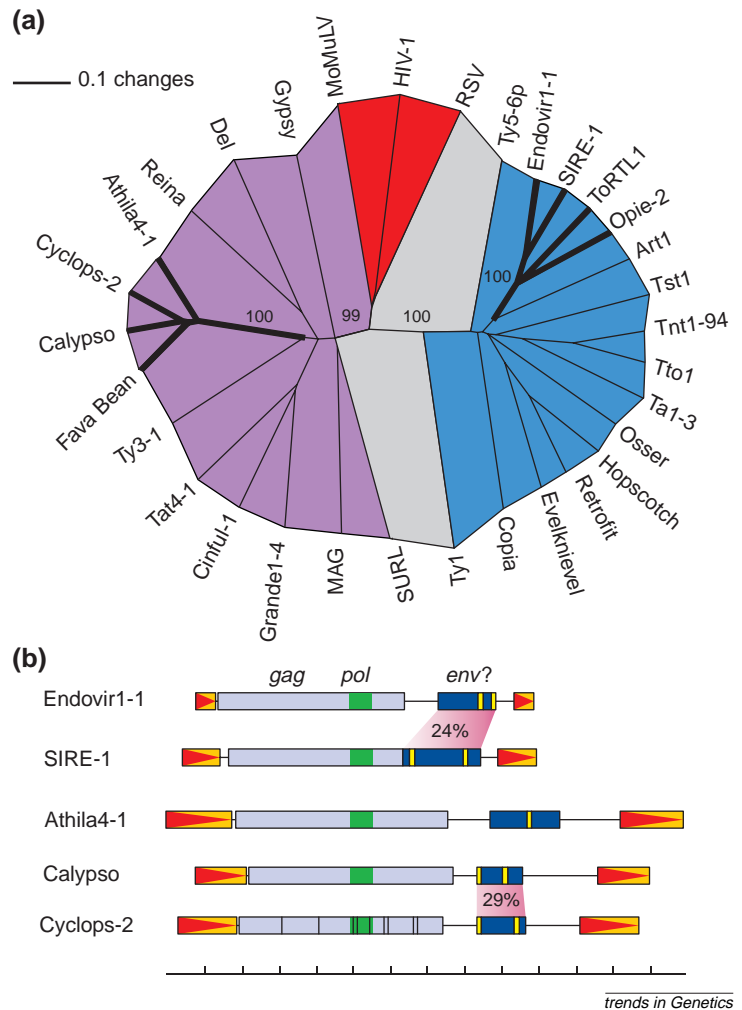
Phylogenetic analysis of reverse transcriptase sequences has divided LTR retroelements into two major lineages^{1,2} – one constitutes the Ty1/*cop* retrotransposons (pseudoviridae) and the other includes both the Ty3/*gypsy* retrotransposons (metaviridae) and the retroviruses (retroviridae) (Fig. 1a). Although retroviruses were first described in vertebrates, a handful of invertebrate Ty3/*gypsy* retroelements possess an additional *env*-like open reading frame (ORF)^{3–5}. At least one of these elements, *gypsy* from *Drosophila melanogaster*, is infectious^{6,7}. The presence of an *env*-like gene that encodes a transmembrane protein is generally considered to be a predictor of a retroelement's infectious nature.

We recently reported plant retroelements with *env*-like genes^{8,9}, implying that enveloped retroviruses are not limited to animals. Candidate plant retroviruses include the *Athila* elements of *Arabidopsis thaliana* and the *SIRE-1* elements of soybean (Fig. 1b). *Athila* and *SIRE-1* have both been successful in colonizing their host's genomes; the *Athila* elements are among the most abundant *A. thaliana* retroelements described to date (0.3% of the genome)¹⁰, and several hundred *SIRE-1* elements reside in soybean¹¹. Interestingly, *SIRE-1* is a member of the Ty1/*cop* lineage, for which infectious retroviruses have not yet been described. The possibility that some Ty1/*cop* elements are infectious suggests at least two independent origins of the retroviruses.

In our initial descriptions of the potential plant retroviruses, we left a critical question unanswered: are the *env*-like genes conserved in retroelements from distantly related plant species, as would be predicted if they were related by descent and played an essential role in the retroelement life cycle? An alternative hypothesis would be that they are derived from transduced gene sequences that do not necessarily confer new functions to the elements. Recent sequence submissions to the DNA databases and DNA sequencing of homologs in other plants appear to have answered this question. The emerging *Arabidopsis* genome sequence harbors a family of *SIRE-1* homologs that we call the *Endovir* family (Fig. 1b). *Endovir1-1* has a conserved *env*-like ORF with predicted transmembrane domains that exhibits 24% amino acid identity to *SIRE-1*. Two *Athila* homologs have also been characterized: the *Cyclops* element of pea¹² and a family of elements from soybean that we report here, named

Calypso. Although the *Calypso* elements are highly degenerate, a consensus sequence based on four independent insertions bears all of the hallmarks of a functional retrovirus. This includes an ORF with predicted transmembrane

FIGURE 1. Comparisons among putative plant retroviruses



(a) Phylogenetic tree showing relationships among the LTR retroelements. The Ty1/*cop* elements (pseudoviridae) are shaded in blue; Ty3/*gypsy* elements (metaviridae) are shaded in purple; retroviruses (retroviridae) are shaded in red. Thick branches represent the *SIRE-1/Endovir* and *Athila/Cyclops* lineages. Reverse transcriptase amino acid sequence alignments are as described in Ref. 2, and they are available at our website, as are additional details of tree construction and sequence analysis (<http://www.public.iastate.edu/~vovtas>). The tree was generated by the neighbor-joining distance algorithm using PAUP^{13,14} and is unrooted. Bootstrap values (1000 replicates) for branchpoints relevant to this study are shown; these branchpoints were also supported by parsimony analysis (data not shown). The *Calypso* sequence used in this analysis and Fig. 1b is a strict consensus sequence of four independent insertions (GenBank accessions AF186182–AF186186). The *Endovir* sequence was identified as part of the *A. thaliana* genome sequencing effort and can be found on P1 MQD19, base position 51655–60737 (GenBank accession AB026651). (b) Structural organization of putative plant retroviruses. Boxes with triangles represent LTRs, and open boxes represent open reading frames. Vertical lines within the boxes indicate frameshifts or stop codons. All elements are aligned by their encoded reverse transcriptase, which is indicated in green. The *env*-like ORFs are dark blue, and putative transmembrane domains are shown in yellow. Transmembrane domains were identified by the TMpred program¹⁵. The scale below the figure is in 1 kb units.

Brooke D. Peterson-Burch
brookepb@iastate.edu

David A. Wright
wrightd@iastate.edu

*Howard M. Laten
hlaten@orion.it.luc.edu

Daniel F. Voytas
voytas@iastate.edu

Department of Zoology and Genetics, 2208 Molecular Biology Building, Iowa State University, Ames, IA 50011, USA.

*Department of Biology, Loyola University Chicago, Chicago, IL 60626, USA.

domains that exhibits 29% amino acid identity to the *env*-like ORF encoded by *Cyclops*. Phylogenetic analyses based on retroelement reverse transcriptases indicate that *SIRE-1/Endovir* elements form a distinct branch among the Ty1/*cop* elements, as do the *Athila/Cyclops* family within the Ty3/*gypsy* lineage (Fig. 1a).

Defining the function of the *env*-like ORFs is the next task at hand. Based on the role of envelope proteins in insect and vertebrate retroviruses, the most obvious possibility is that they interact with host cell membranes to enable infection. Although the plant cell wall presents a barrier to membrane-mediated transmission, enveloped viruses do exist in plants. Some rhabdoviruses and bunyaviruses possess genes that encode envelope glycoproteins, which enable these viruses to shuttle between invertebrate and plant hosts¹. It is possible that the putative plant retroviruses also use insect hosts as agents for transmission. At least one of the *Arabidopsis Endovir* insertions is structurally intact, and a potentially functional *Athila* element can be generated from the consensus sequence of characterized insertions (Fig. 1b). This suggests that the role of the *env*-like ORF can now be tested by developing assays

for transposition and infection. Should they prove infectious, it might be that retroviruses, because of their ability to occasionally transduce cellular genes, are a potent vehicle for interspecies gene flow in plants. Unlike animals, plants do not sequester their germ line, and infected somatic plant cells can give rise to floral organs and seeds. Therefore, over evolutionary timescales, plant genomes could have been greatly influenced by retroviral infections.

Plant retroviruses also hold much promise for the deliberate genetic modification of plants. Disarmed, non-infectious animal retroviruses are widely used as vectors for gene transfer in vertebrates. A similar plant retroviral vector system could be developed to help harness the natural biosynthetic capacity of plants for the production of food, pharmaceuticals and industrial raw materials, an area of burgeoning interest to the plant biotechnology sector. However, before they are used as vectors, it will be important to understand how plant retroviruses naturally contribute to interspecies gene flow. This information will be necessary to rationally evaluate recent concerns regarding the use of genetically modified crop species.

References

<p>1 van Regenmortel, M. <i>et al.</i> (1999) <i>Virus Taxonomy</i>, Academic Press</p> <p>2 Xiong, Y. and Eickbush, T.H. (1990) Origin and evolution of retroelements based upon their reverse transcriptase sequences. <i>EMBO J.</i> 9, 3353–3362</p> <p>3 Britten, R.J. (1995) Active <i>gypsy</i>/Ty3 retrotransposons or retroviruses in <i>Caenorhabditis elegans</i>. <i>Proc. Natl. Acad. Sci. U. S. A.</i> 92, 599–601</p> <p>4 Leblanc, P. <i>et al.</i> (1997) Invertebrate retroviruses: ZAM a new candidate in <i>D. melanogaster</i>. <i>EMBO J.</i> 16, 7521–7531</p> <p>5 Tanda, S. <i>et al.</i> (1994) The <i>Drosophila</i> tom retrotransposon encodes an envelope protein. <i>Mol. Cell. Biol.</i> 14, 5392–5401</p> <p>6 Kim, A. <i>et al.</i> (1994) Retroviruses in invertebrates: the <i>gypsy</i> retrotransposon is apparently an infectious retrovirus of</p>	<p><i>Drosophila melanogaster</i>. <i>Proc. Natl. Acad. Sci. U. S. A.</i> 91, 1285–1289</p> <p>7 Song, S.U. <i>et al.</i> (1994) An <i>env</i>-like protein encoded by a <i>Drosophila</i> retroelement: evidence that <i>gypsy</i> is an infectious retrovirus. <i>Genes Dev.</i> 8, 2046–2057</p> <p>8 Laten, H.M. <i>et al.</i> (1998) <i>SIRE-1</i>, a <i>cop</i>ia/Ty1-like retroelement from soybean, encodes a retroviral envelope-like protein. <i>Proc. Natl. Acad. Sci. U. S. A.</i> 95, 6897–6902</p> <p>9 Wright, D.A. and Voytas, D.F. (1998) Potential retroviruses in plants: <i>Tat1</i> is related to a group of <i>Arabidopsis thaliana</i> Ty3/<i>gypsy</i> retrotransposons that encode envelope-like proteins. <i>Genetics</i> 149, 703–715</p> <p>10 Pelissier, T. <i>et al.</i> (1996) <i>Athila</i>, a new retroelement from <i>Arabidopsis thaliana</i>. <i>Plant Mol. Biol.</i> 29, 441–452</p> <p>11 Laten, H. and Morris, R.O. (1993) <i>SIRE-1</i>, a long interspersed</p>	<p>repetitive DNA element from soybean with weak sequence similarity to retrotransposons: initial characterization and partial sequence. <i>Gene</i> 133, 153–159</p> <p>12 Chavanne, F. <i>et al.</i> (1998) Structure and evolution of <i>Cyclops</i>: a novel giant retrotransposon of the Ty3/<i>gypsy</i> family highly amplified in pea and other legume species. <i>Plant Mol. Biol.</i> 37, 363–375</p> <p>13 Saitou, N. and Nei, M. (1987) The neighbor-joining method: a new method for reconstructing phylogenetic trees. <i>Mol. Biol. Evol.</i> 4, 406–425</p> <p>14 Swofford, D. (1999) <i>PAUP 4.0</i>. Laboratory of Molecular Systematics, Smithsonian Institution</p> <p>15 Hofmann, K. and Stoffel, W. (1993) TMbase – a database of membrane spanning proteins segments. <i>Biol. Chem. Hoppe-Seyler</i> 347, 166</p>
--	--	---

Reverse gyrase from hyperthermophiles probable transfer of a thermoadaptation trait from Archaea to Bacteria

Nelson *et al.*, who recently sequenced the complete genome of the hyperthermophilic bacterium *Thermotoga maritima*, have reported that 24% of its putative open reading frames are more similar to archaeal genes than to other bacterial genes¹. This is a much higher percentage than that found in mesophilic bacteria, which show similarity to archaeal genes in the range of 3–7%, and suggested that extensive lateral gene transfer has occurred between archaea and *T. maritima*. Similar observations and conclusions have previously been drawn by Koonin and coworkers, who have also noticed a high percentage of archaeal genes (15%) in the hyperthermophilic bacterium *Aquifex aeolicus*². As an alternative to gene transfer, it was suggested that these genes were vertically

inherited from the last hyperthermophilic ancestor of Archaea and Bacteria, and lost in mesophilic bacteria³. However, many specific archaeal genes present in *T. maritima* and *A. aeolicus* are grouped in clusters^{1,2}, supporting the hypothesis of gene transfer (other arguments in favour of transfer are discussed in Ref. 4). A major question is how to determine if these transfers are related to the adaptation of these bacteria to life at high temperature, and if they have occurred mainly from archaea to hyperthermophilic bacteria or *vice versa*.

Nelson *et al.*¹ list 108 genes from *T. maritima* that match only with genes from hyperthermophiles. They missed among them the gene encoding *T. maritima* reverse gyrase (*rgy*) (Tma0173). This is probably because,