



Phylogenetic evidence for *Ty1-copia*-like endogenous retroviruses in plant genomes

Howard M. Laten

Biology Department, Loyola University Chicago, 6525 N. Sheridan Rd., Chicago, IL 60626 (Phone: 1-773-508-3640; Fax: 1-773-508-3646; E-mail: hlaten@luc.edu)

Accepted 31 January 2000

Key words: *Arabidopsis*, envelope, retrotransposon, reverse transcriptase, soybean

Abstract

SIRE-1 is a multi-copy, *Ty1-copia*-like retroelement family found in the genome of *Glycine max*. A sequenced *SIRE-1* genomic copy has an uninterrupted ORF that can be translated into a gag-pol polyprotein, followed by an unprecedented second ORF whose conceptual translation yielded a theoretical protein predicted to possess many of the same secondary structural elements found in mammalian retroviral envelope proteins. Similar, but clearly pseudogenic, envelope-like sequences were recovered from conceptual translations of 10 *Arabidopsis* GenBank accessions. All were associated with identifiable *Ty1-copia*-like retroelements. Phylogenetic analysis of the adjacent ribonuclease H regions from these sequences and three similarly endowed elements, two from maize and one from tomato, indicate that the 14 elements constitute a monophyletic group distinct from several closely related plant *Ty1-copia*-like elements in which *pol* is immediately followed by a downstream LTR. The conservation of identifiable *env*-like gene features suggests that these plant elements are endogenous retroviruses whose ancestors were acquired from animal vectors. The finding that the *env* and *env*-less retroelements identified in this study form distinct lineages does not support the hypothesis that horizontal transmission of retrotransposons is sponsored by ancestral infectious retroviruses that subsequently lost all traces of *env* genes.

Abbreviations: LTR – long terminal repeat; RNAase H – ribonuclease H; ORF – open reading frame; MP – movement protein.

Introduction

Retroelements are ubiquitous components of bacterial and eukaryotic genomes that employ reverse transcriptase to sponsor their proliferation (Eickbush, 1994; Boeke & Stoye, 1997). They encompass a diverse collection of genetic elements that include DNA and RNA viruses (Petropoulos, 1997; Eickbush, 1994), endogenous retroviruses (Boeke & Stoye, 1997; Patience, Wilkenson & Weiss, 1997), retrotransposons (Boeke & Stoye, 1997; Eickbush, 1994), long (LINEs) and short (SINEs) interspersed nuclear elements (Boeke & Stoye, 1997; Eickbush, 1994), and *Drosophila* telomeres (Pardue et al., 1996). Long terminal repeat (LTR) retrotransposons and LINEs con-

stitute as much as 50% of some eukaryotic genomes (Bennetzen, 1996; Eickbush, 1994).

Infectious and endogenous retroviruses and related, non-infectious retrotransposons are distinguished from other retroelements, including LINEs, by their possession of LTR's (Figure 1, Boeke & Stoye, 1997; Eickbush, 1994). LTR retrotransposons have been identified in the genomes of several vertebrates (Britten et al., 1995; Flavell et al., 1995; Flavell & Smith, 1992), and are routinely found in the genomes of lower animals, plants, and fungi (Boeke & Stoye, 1997; Britten, 1995; Flavell, Smith & Kumar, 1992; Kumar, 1996; Suoniemi, Tanskanen & Schulman, 1998; Voytas et al., 1992). The presence of *env* genes (or pseudogenes) encoding envelope glycopro-

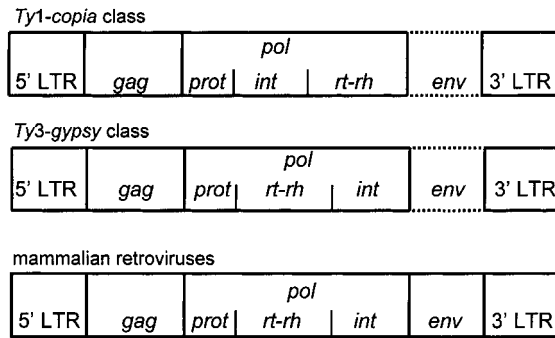


Figure 1. Organization of LTR retroelements. *gag*: core structural protein genes; *prot*: aspartic acid protease; *int*: integrase; *rt-rh*: reverse transcriptase RNAase H; *env*: envelope. Dotted lines indicate that some members of group have putative *env*.

teins that facilitate cell to cell viral transfer or *env*-like open reading frames (ORF's) is the major feature that distinguishes infectious and endogenous retroviruses from LTR retrotransposons. Infectious retroviruses are primarily associated with mammals and birds, but are present in other vertebrates (Petropoulos, 1997). Endogenous retroviruses are present in most vertebrate genomes (Boeke & Stoye, 1997) and the genomes of a few invertebrates (Britten, 1995; Felder et al., 1994; Inouye, Yuke & Saigo, 1986; Kim et al., 1994; Leblanc et al., 1997; Ozers & Friesen, 1996; Saigo et al., 1984; Song et al., 1994; Tanda, Mullor & Corces, 1994). In insects, these endogenous retroviruses have been designated errantiviruses (Boeke et al., 1999). Based on the criterion of possession of *env*-like ORF's, endogenous retroviruses are arguably also present in the genomes of higher plants (Laten, Majumdar & Gaucher, 1998; Wright & Voytas, 1998; Kumar, 1998; Peterson-Burch et al., 2000).

Most of the characterized LTR retrotransposons belong to either the *Ty1-copia* or *Ty3-gypsy* group (Boeke & Stoye, 1997; Eickbush, 1994). The two classes can be phylogenetically segregated by amino acid comparisons of the reverse transcriptase domain (Eickbush, 1994; Doolittle & Feng, 1992) and by the order of the catalytic domains in *pol* (Figure 1). All vertebrate infectious and endogenous retroviruses conform to the *Ty3-gypsy* configuration (Eickbush, 1994), and phylogenetic analyses of the conserved regions within the reverse transcriptase suggest that vertebrate retroviruses and *Ty3-gypsy* retroelements may be monophyletic (Eickbush, 1994).

SIRE-1 is a moderately high copy-number, *Ty1-copia*-like LTR retroelement family that resides in the genomes of cultivated *Glycine max*, its wild progen-



Figure 2. Organization of *SIRE-1* ORF's. See Figure 1 for abbreviations.

itor, *Glycine soja*, and other wild members of the genus (Bi & Laten, 1996; Laten & Morris, 1993; Laten, unpublished). Although direct evidence for functional expression has not been generated, *SIRE-1* sequences are well represented in cDNA clones (Bi & Laten, 1996), and cDNA's encoding nucleocapsid, reverse transcriptase, and envelope domains have been recovered by RT-PCR (Lin & Laten, unpublished). A sequenced *SIRE-1* genomic copy has an uninterrupted ORF of 1552 codons that can be translated into a theoretical polypeptide with conserved nucleocapsid, aspartic protease, integrase, reverse transcriptase and ribonuclease H (RNAase H) domains (Figure 2, Laten, Majumdar & Gaucher, 1998; Laten, Gaucher & Majumdar, unpublished). The RNAase H coding region at the 3' end of *pol* is followed by a second ORF of 648 codons (Laten, Majumdar & Gaucher, 1998). Conceptual translation of ORF2 generated a protein with a molecular weight of 70 kD. This theoretical protein is predicted to possess many of the same secondary structural elements found in mammalian retroviral envelope proteins, including trans-membrane domains, a coiled coil, and multiple, lengthy α helices (Laten, Majumdar & Gaucher, 1998).

Methods

TBLASTN searches (Altschul et al., 1997) of the non-redundant DNA database at NCBI were run initially using the conceptual translation of ORF2 from *SIRE-1* as query sequence. Sequences recovered from this search were then used as the queries for subsequent TBLASTN searches.

The nucleotide sequences between *pol* and the 3'-LTR of all plant *Ty1/copia*-like elements recovered in this paper or previously reported were translated, if necessary, into three forward reading frames (relative to *pol*). The sequences and their sources are listed in Table 1. Where necessary, frameshifts were corrected manually to conform to consensus sequences, and stop codons were deleted. The sequences were then aligned pairwise using LALIGN (Biology Workbench 3.0, National Center for Supercomputing Applica-

tions, University of Illinois at Urbana-Champaign, 1998) to evaluate the location(s) and extent of amino acid identities and similarities.

The reverse transcriptase regions and RNAase H regions of each element were recovered, translated, and aligned using PILEUP (Wisconsin Package Version 9.1, Genetics Computer Group, Madison, USA). PAUPSEARCH (PAUP* 4.0) was used to generate bootstrapped maximum parsimony and neighbor joining cladograms.

Results and discussion

Recovery and alignment of env-like sequences from Ty1-copia-like elements

Using the theoretical translation of the *env*-like gene from *SIRE-1* as the query sequence, TBLASTN searches were run to recover related sequences from the collective, non-redundant DNA database at NCBI. The initial search recovered seven sequences with E values below 10, ranging from $5e^{-16}$ to 2.4, all from *Arabidopsis thaliana*. None of the recovered sequences with E values between 10 and 100 were plant DNA's, nor were any related to retroelements. However, three additional *A. thaliana* sequences were recovered when the candidate sequences from the first run were used as queries in subsequent TBLASTN searches. Extensive deletions in these sequences were the cause of their initial omission.

The 10 sequences were evaluated to ascertain the identity of flanking regions. In all cases, the sequences were clearly associated with identifiable, *Ty1-copia*-like retrotransposons. Reverse transcriptase-like sequences were located immediately upstream (relative to the putative coding strand) of all 10 sequences, and one member of a pair of LTR's was located immediately downstream of all but one sequence. The exception, missing only a 3' LTR, abutted the LTR of an adjacent, *Ty3-gypsy*-like retrotransposon.

Generating useful amino acid alignments from the envelope-like sequences to construct a meaningful phylogenetic tree was hindered by the presence of multiple deletions and the sparsity of informative sites. All but the *SIRE-1* sequence were clearly pseudo-genes. However, the relatedness of the recovered sequences was evaluated by pairwise comparisons using LALIGN (Table 2). The analysis included three additional *Ty1-copia*-like retrotransposons, *Opie-2*,

PREM-2, and ToRTL1, that appear to contain pseudo-gene regions between *pol* and the 3' LTR. However, while these regions in *Opie-2* and PREM-2 are closely related, neither of them generated a significant alignment with any other sequence. AB016882, AC005970, and AF076243 share a deletion that limits their overlap with the remaining elements to 30–60 amino acids, and are not included in Table 2. Figure 3 illustrates the overlapping regions of significant amino acid identity. Table 2 and Figure 3 support the argument that all of the sequences are related and do not have independent origins. That the relatedness of a retroelement gene outside of *pol* can be clearly demonstrated by amino acid alignment is remarkable, especially considering the wide range of host orders: Sapindales, Brassicales and Solanales. This relatively strong conservation, and the absence of frameshift and nonsense codons in one member, *SIRE-1*, suggests that the gene has been selectively retained and has persisted for 10s of millions of years, horizontal transfer notwithstanding.

RNAase H-based phylogenetic relationships of plant Ty1-copia-like retroelements

The reverse transcriptase and RNAase H regions immediately upstream of the putative *env* genes were used as queries for TBLASTN searches to identify a set of retrotransposon-like elements most closely related to those collected above, independent of the presence of the envelope-like coding region. Sequences of well characterized *Ty1-copia*-like retrotransposons were also included. The sequences were aligned using PILEUP from the Wisconsin Package, but multiple, lengthy deletions distributed throughout the reverse

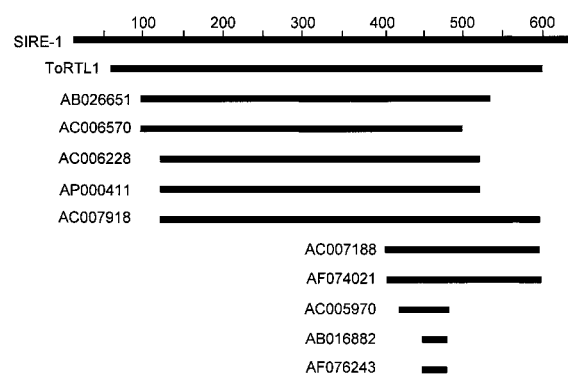


Figure 3. Regions of amino acid identities within envelope-like theoretical proteins of *Ty1-copia* like elements aligned to the *env*-like ORF of *SIRE-1*. Lengths of identity may not correspond to Table 2 because of deletions and insertions.

Table 1. Source of DNA sequences used in this analysis. A. Previously characterized *Ty1-copia* retrotransposons. B. Unpublished *A. thaliana* sequences from GenBank

Element	A		B	
	Accession	Organism	Accession	Base range ^a
SIRE-1	U96295	<i>G. max</i>	AB006567	81525–81932c
ToRTL1	U68072	<i>L. esculentum</i>	AB016882	16076–17717c
Opie-2	U68408	<i>Z. mays</i>	AB019224	58296–58706
PREM-2	U41000	<i>Z. mays</i>	AB026651	56799–59464
Tgmr	U96748	<i>G. max</i>	AC005970	61334–63961
Tnt1	X13777	<i>N. tabacum</i>	AC006228	5460–7944
Tto1	D83003	<i>N. tabacum</i>	AC006570	3833–6012
Tpv2-18	AJ005763	<i>P. vulgaris</i>	AC007188	62690–65100c
Hopscotch	U12626	<i>Z. mays</i>	AC007918	4600–6829
Retrofit	U72725	<i>O. longistamata</i>	AC012392	123023–123430
PSretro	AJ243358	<i>P. sativum</i>	AF074021	42023–44427c
Ta1-1	X53973	<i>A. thaliana</i>	AF076243	20140–22201c
LB3RT	AJ001752	<i>V. faba</i>	AP000411	53259–55774c
Copia	X04456	<i>D. melanogaster</i>		

^ac-complementary strand.

Table 2. Amino acid identities among *env* regions of *Ty1-copia* retroelements generated using LALIGN^a

	ToRTL1	AP000411	AC006228	AC006570	AC007188	AC007918	AB026651	AF074021
SIRE-1	20.3	25.2	27.3	29.7	26.5	27.1	25.6	27.0
	582	325	311	233	136	269	446	137
ToRTL1	–	NS	21.7	NS	NS	NS	34.1	NS
			300				82	
AP000411	–	–	61.7	27.0	24.1	32.3	27.9	30.8
			522	222	145	164	305	143
AC006228	–	–	–	29.2	24.5	29.0	29.9	25.0
				178	249	241	338	336
AC006570	–	–	–	–	42.9	30.4	27.1	43.4
					105	313	299	106
AC007188	–	–	–	–	–	47.4	31.2	65.3
						211	141	623
AC007918	–	–	–	–	–	–	32.6	61.1
							184	131
AB026651	–	–	–	–	–	–	–	26.1
								119

^aTop line of each entry is the % amino acid identity; bottom line is the number of contiguous amino acids over which the identities were determined. NS, not significant: LALIGN raw scores of less than 100 were considered not significant.

transcriptase domains of many of the *Arabidopsis* elements limited useful sequence comparisons to the last 130 amino acids of RNAase H. Using PAUP*, the aligned RNAase H sequences generated a unique, most parsimonious tree (Figure 4). The tree shows that the entire collection of known *Ty1-copia*-like retroelements containing an *env*-like ORF or pseudogene,

including the two from maize that did not generate significant envelope alignments, is monophyletic. A bootstrapped, neighbor joining tree had similar topology (not shown). The tree is rooted using *copia* as the outgroup.

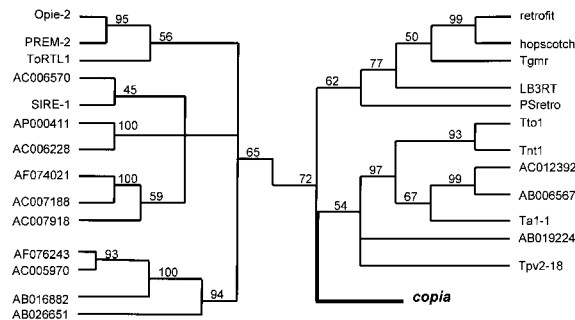


Figure 4. Maximum parsimony tree generated using PAUP* of selected *Ty1-copia* group retroelements. Numbers adjacent to nodes are bootstrap values for 100 replicates. Only nodes supported by at least 45% of resamplings are shown. Root shown in bold with *copia* as outgroup. See Table 1 for references.

Plant endogenous retroviruses?

A subset of plant retrotransposons contains ORF's or pseudogenes whose conceptual translation yields proteins that resemble retroviral envelope proteins (Laten, Majumdar & Gaucher, 1998; Wright & Voytas, 1998; this paper), and we and others have proposed that they may constitute endogenous retroviruses (Kumar, 1998). These elements include members from both *Ty3-gypsy* (Wright & Voytas, 1998; Peterson-Burch et al., 2000) and *Ty1-copia* (Laten, Majumdar & Gaucher, 1998; this paper) families. Endogenous retroviruses belonging to the *Ty3-copia* class have also been characterized in several invertebrate species (Boeke & Stoye, 1997).

The possibility that the putative envelope-like proteins in LTR retrotransposons are actually transduced genes unrelated to infectious transfer has been considered. The conservation of predicted structural features of the theoretical envelope-like proteins for invertebrate and plant *Ty3-gypsy* elements (Boeke & Stoye, 1997; Wright & Voytas, 1998; Peterson-Burch et al., 2000) and the plant *Ty1-copia* elements (Laten, Majumdar & Gaucher, 1998; this paper) suggests that they are functionally analogous to each other and maintained by selection. It is unlikely that these similarities are simply coincidental, and infectious transfer is the only function with a clear precedent, that of mammalian retroviruses. And while retroviral transduction of host genes is well documented, the only LTR retrotransposon that appears to have transduced a host gene, *Bs1* from maize, has no apparent descendants, and has forfeited its *pol* gene and mobility to acquire an ATPase cDNA (Bureau, White & Wessler, 1994; Jin & Bennetzen, 1994).

The proposal that plant genomes harbor proretroviruses that encode trans-membrane proteins with viral envelope function evokes two perplexing concerns. First, envelope-mediated viral infection is an animal virus strategy that would be compromised by plant cell walls. Second, no genes encoding proteins that sponsor cell-to-cell or long distance viral transport in plant hosts (Carrington et al., 1996; Ding, 1998) have been identified in any retroelement sequence.

The presence of envelope genes in plant viruses is not unusual. Plant viral genomes that encode envelope proteins – bunyaviruses and some rhabdoviruses (Brunt et al., 1997) – also mount persistent infections in animal hosts. *Env* genes are apparently expendable in the plant host, although they are expressed. When tospoviruses, plant members of the Bunyaviridae, are maintained solely by mechanical inoculation of host plants, morphologically defective isolates can be recovered with no selection that carry disabling point and frameshift mutations in *env* (Goldbach & Peters, 1996). These isolates are fully infectious in the plant but not in the invertebrate host. On the other hand, functional *env* genes are required for propagation in, and possibly transfer to, invertebrate hosts like thrips, planthoppers, and aphids, which presumably ingest enveloped viral particles during feeding (Goldbach & Peters, 1996; Nault, 1994). The family of endogenous plant elements from *Arabidopsis* and maize on the 'enveloped' arm of Figure 4, may be vestiges of similarly transmitted retroviruses. The nearly intact *env*-like ORF's in *ToRTL1* and *SIRE-1* suggest that the integrity of these sequences has been maintained by relatively recent selective pressure, presumably in an invertebrate host.

Movement proteins (MP) are a highly diverse collection of small proteins (25–35 kD) that are required for the intra- and inter-cellular transport of virions or viral genomes in higher plants (Carrington et al., 1996). Transfer is accomplished by MP association with capsids, nucleocapsids or viral genomes, and MP-sponsored trafficking through plasmodesmata or intercellular tubules (Carrington et al., 1996). *NSm*, the MP gene in the tospovirus, TSWV, is not expressed in its invertebrate host (Mumford, Barker & Wood, 1996), and animal bunyaviruses lack this gene entirely (Goldbach & Peters, 1996). There is a third ORF in *SIRE-1* (Figure 2) whose conceptual translation produces a 22 kD peptide that can be extended to 28 kD with readthrough of a single stop codon. This ORF is located within the 5' half of the *env*-like ORF, but in the antisense orientation. Although the amino acid

sequence provides little information about the possible function of this theoretical *SIRE-1* peptide, the ambisense coding strategy is used by several plant viruses, including tospoviruses, to express MP genes (Goldbach & Peters, 1996; Mumford, Barker & Wood, 1996).

Possible multiple origins of retroviruses and horizontal transfer of retrotransposons

The possibility that in addition to the *Ty3-gypsy* group, some *Ty1-copia* members may actually be endogenous or infectious retroviruses, suggests that retroviruses evolved at least twice. The observation also provides a foundation for the proposal that the apparent horizontal transfer of LTR retrotransposons may be the result of transmission of ancestral retroviral derivatives that subsequently lost their extracellular transmission genes. Retroviruses are presumed to have evolved from retrotransposons through acquisition of an *env* gene (Doolittle & Feng, 1992; Eickbush, 1994; Temin, 1992), and while most published phylogenetic trees reflect this supposition (Doolittle & Feng, 1992; Eickbush, 1994), the reverse relationship is not inconsistent with the data and arguments for the reverse scenario have been made (Coffin, 1993; Lerat & Capy, 1999). Infectious transfer of a retroviral form of existing retrotransposons could account for the apparent interspecies spread of the latter if deletion of *env* was frequent in certain hosts and/or if its absence enhanced intracellular retrotransposition of the newly introduced element. Since retroelement replication proceeds through an RNA intermediate, and the expression of the retrovirus *env* gene involves a spliced transcript (Eickbush, 1994), it is not unreasonable to envision reverse transcriptase sometimes utilizing a misspliced RNA lacking *env* as template for the synthesis and subsequent integration of a non-infectious, retrotransposable DNA. This model would account for both horizontal transfer and subsequent vertical spread of non-infectious elements. However, for the *SIRE-1*-related sequences currently available, the tree shown in Figure 4 does not support this hypothesis since the *env* and *env*-less groups are each monophyletic. Wright and Voytas (1998) found an analogous relationship among a group of plant *Ty3-gypsy*-like retroelements with an *env*-like feature, and the same has been inferred for errantiviruses (Lerat & Capy, 1999). As additional genomic sequences become available, the question of origins should be re-evaluated.

Acknowledgements

The author wishes to acknowledge the contributions of Melissa Sopko, Jaqui Wittke, Dawn Wilson, and Melissa DeRosa, for help with retrieving sequences from the database, Drs. Eric Brown and Terry Grande for help with phylogenetic tools and interpretations, and Dr. Dan Voytas and David Wright for sharing unpublished data.

References

- Altschul, S.F., T.L. Madden, A.A. Schäffer, J. Zhang, Z. Zhang, W. Miller & D.J. Lipman, 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* 25: 3389–3402.
- Bennetzen, J.L., 1996. The contributions of retroelements to plant genome organization, function and evolution. *Trends Microbiol.* 4: 347–353.
- Bhattacharyya, M.K., R.A. Gonzales, M. Kraft & R.I. Buzzell, 1997. A *copia*-like retrotransposon *Tgmr* closely linked to the *Rps-k* allele that confers race-specific resistance of soybean to *Phytophthora sojae*. *Plant Mol. Biol.* 34: 255–264.
- Bi, Y.-A. & H.M. Laten, 1996. Sequence analysis of a cDNA containing the *gag* and *prot* regions of the soybean retrovirus-like element, *SIRE-1*. *Plant Mol. Biol.* 30: 1315–1319.
- Boeke, J.D., T. Eickbush, S.B. Sandmeyer & D.F. Voytas, 1999. Metaviridae, In *Virus Taxonomy: ICTV VIIth Report*, edited by M.H.V. van Regenmortel, C.M. Fauquet, D.H.L. Bishop, E.B. Carstens, M.K. Estes, S.M. Lemon, J. Maniloff, M.S. Mayo, D.J. McGeoch, C.R. Pringle & R.B. Wickner. Academic Press, NY, in press.
- Boeke, J.D. & J.P. Stoye, 1997. Retrotransposons, endogenous retroviruses, and the evolution of retroelements, pp. 343–436 in *Retroviruses*, edited by J.M. Coffin, S.H. Hughes & H.E. Varmus. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
- Britten, R.J., 1995. Active *gypsy/Ty3* retrotransposons or retroviruses in *Caenorhabditis elegans*. *Proc. Natl. Acad. Sci. USA* 92: 599–601.
- Britten, R.J., T.J. McCormack, T.L. Mears & E.H. Davidson, 1995. *Gypsy/Ty3*-class retrotransposons integrated in the DNA of herring, tunicate, and echinoderms. *J. Mol. Evol.* 40: 13–24.
- Brunt, A.A., K. Crabtree, M.J. Dallwitz, A.J. Gibbs, L. Watson & E.J. Zurcher (eds.), 1997. *Plant Viruses Online: Descriptions and Lists from the VIDE Database, Version 16*.
- Bureau, T.E., S.E. White & S.R. Wessler, 1994. Transduction of a cellular gene by a plant retroelement. *Cell* 77: 479–480.
- Carrington, J.C., K.D. Kasschau, S.K. Mahajan & M.C. Schaad, 1996. Cell-to-cell and long-distance transport of viruses in plants. *Plant Cell* 8: 1669–1681.
- Coffin, J.M., 1993. Reverse transcriptase and evolution, pp. 445–479 in *Reverse Transcriptase*, edited by A.M. Skalka & S.P. Goff. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
- Ding, B., 1998. Intercellular protein trafficking through plasmodesmata. *Plant Mol. Biol.* 38: 279–310.
- Doolittle, R.F. & D.-F. Feng, 1992. Tracing the origin of retroviruses. *Curr. Top. Microbiol. Immunol.* 176: 195–211.
- Eickbush, T.H., 1994. Origin and evolutionary relationships of retroelements, pp. 121–157 in *The Evolutionary Biology of Viruses*, edited by S.S. Morse. Raven Press, NY.

- Felder, H., A. Herzceg, Y. deChastonay, P. Aeby, H. Tobler & F. Muller, 1994. *Tas*, a retrotransposon from the parasitic nematode *Ascaris lumbricoides*. *Gene* 149: 219–225.
- Flavell, A.J., V. Jackson, M.P. Iqbal, I. Riach & S. Waddell, 1995. *Ty1-copia* group retrotransposon sequences in amphibia and reptilia. *Mol. Gen. Genet.* 246: 65–71.
- Flavell, A.J. & D.B. Smith, 1992. A *Ty1-copia* group retrotransposon sequence in a vertebrate. *Mol. Gen. Genet.* 233: 322–326.
- Flavell, A.J., D.B. Smith & A. Kumar, 1992. Extreme heterogeneity of *Ty1-copia* group retrotransposons in plants. *Mol. Gen. Genet.* 231: 233–242.
- Goldbach, R. & D. Peters, 1996. Molecular and biological aspects of tospoviruses, pp. 129–157 in *The Bunyaviridae*, edited by R.M. Elliott. Plenum Press, NY.
- Inouye, S., S. Yuki & K. Saigo, 1986. Complete nucleotide sequence and genome organization of a *Drosophila* transposable genetic element, 297. *Eur. J. Biochem.* 154: 417–425.
- Jin, Y.-K. & J.L. Bennetzen, 1994. Integration and nonrandom mutation of a plasma membrane proton ATPase gene fragment within the *Bs1* retroelement in maize. *Plant Cell* 6: 1177–1186.
- Kim, A., C. Terzian, P. Santamaria, A. Pelisson, N. Prud'homme & A. Bucheton, 1994. Retroviruses in invertebrates: the *gypsy* retrotransposon is apparently an infectious retrovirus of *Drosophila melanogaster*. *Proc. Natl. Acad. Sci. USA* 91: 1285–1289.
- Kumar, A., 1996. The adventures of the *Ty1-copia* family of retrotransposons in plants. *Trends Genet.* 12: 41–43.
- Kumar, A., 1998. The evolution of plant retroviruses: moving to greener pastures. *Trends Plant Sci.* 3: 371–374.
- Laten, H.M. & R.O. Morris, 1993. *SIRE-1*, a long interspersed repetitive DNA element from soybean with weak sequence similarity to retrotransposons: initial characterization and partial sequence. *Gene* 133: 153–159.
- Laten, H.M., A. Majumdar & E.A. Gaucher, 1998. *SIRE-1*, a *copia/Ty1*-like retroelement from soybean, encodes a retroviral envelope-like protein. *Proc. Natl. Acad. Sci. USA* 95: 6897–6902.
- Leblanc, P., S. Deswset, B. Dastugue & C. Vaury, 1997. Invertebrate retroviruses – *Zam*, a new candidate in *D. melanogaster*. *EMBO J.* 16: 7521–7531.
- Lerat, E. & P. Capy, 1999. Retrotransposons and retroviruses: Analysis of the envelope gene. *Mol. Biol. Evol.* 16: 1198–1207.
- Mumford, R.A., I. Barker & K.R. Wood, 1996. The biology of tospoviruses. *Ann. Appl. Biol.* 128: 159–183.
- Nault, L.R., 1994. Transmission biology, vector specificity and evolution of planthopper-transmitted plant viruses, pp.429–448 in *Planthoppers: Their Ecology and Management*, edited by R.F. Denno & T.J. Perfect. Chapman and Hall, NY.
- Ozers, M.S. & P.D. Friesen, 1996. The *Emv*-like open reading frame of the baculovirus-integrated retrotransposon TED encodes a retrovirus-like envelope protein. *Virology* 226: 252–259.
- Pardue, M.L., N. Danilevskaya, K. Lowenhaupt, F. Slot & K.L. Tra-verse, 1996. *Drosophila* telomeres: new views on chromosome evolution. *Trends Genet.* 12: 48–52.
- Patience, C., D.A. Wilkenson & R.A. Weiss, 1997. Our retroviral heritage. *Trends Genet.* 13: 116–120.
- Peterson-Burch, B.D., D.A. Wright, H.M. Laten & D.F. Voytas, 2000. Retroviruses in Plants? *Trends Genet.* 16: 151–152.
- Petropoulos, C., 1997. Retroviral taxonomy, protein structures, sequences, and genetic maps, pp. 757–805 in *Retroviruses*, edited by J.M. Coffin, S.H. Hughes & H.E. Varmus, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
- Saigo, K., W. Kugimiya, Y. Matsuo, S. Inouye, K. Yoshioka & S. Yuki, 1984. Identification of the coding sequence for a reverse transcriptase-like enzyme in a transposable genetic element in *Drosophila melanogaster*. *Nature* 312: 659–661.
- SanMiguel, P., A. Tikhonov, Y.-K. Jin, N. Motchoulskaia, D. Zakharov, A. Melake-Berhan, P.S. Springer, K.J. Edwards, M. Lee, Z. Avramova & J.L. Bennetzen, 1996. Nested retrotransposons in the intergenic regions of the maize genome. *Science* 274: 765–768.
- Song, S.U., T. Gerasimova, M. Kurkulos, J.D. Boeke & V.G. Corces, 1994. An *Env*-like protein encoded by a *Drosophila* retroelement: evidence that *gypsy* is an infectious retrovirus. *Genes Dev.* 8: 2046–2057.
- Suoniemi, A., J. Tanskanen & A.H. Schulman, 1998. *Gypsy*-like retrotransposons are widespread in the plant kingdom. *Plant J.* 13, 699–705.
- Tanda, S., J.L. Mullor & V.G. Corces, 1994. The *Drosophila tom* retrotransposon encodes an envelope protein. *Mol. Cell. Biol.* 14: 5392–5401.
- Temin, H.M., 1992. Origin and general nature of retroviruses, pp. 1–18 in *The Retroviridae*, Vol 1, edited by J.A. Levy. Plenum Press, NY.
- Voytas, D.F., M.P. Cummings, A. Konieczny, F.M. Ausubel & S.R. Rodermel, 1992. *copia*-like retrotransposons are ubiquitous among plants. *Proc. Natl. Acad. Sci. USA* 89: 7124–7128.
- Wright, D.A. & D.F. Voytas, 1998. Potential retroviruses in plants: *Tat1* is related to a group of *Arabidopsis thaliana* *Ty3/gypsy* retrotransposons that encode envelope-like proteins. *Genetics* 149: 703–715.