Review of Literature
THE FAMILY OF RETROELEMENTS:

The family of retroelements is composed of a variety of genetic elements that contain a gene for the enzyme reverse transcriptase, which catalyzes the synthesis of DNA molecule from an RNA template. The RNA dependent DNA polymerase or reverse transcriptase was first discovered nearly twenty five years ago as a retroviral encoded enzyme catalyzing DNA synthesis from RNA template (Baltimore, 1970; Temin and Mizutami, 1970). Since then, many genetic elements from a wide variety of organisms have been shown to contain open reading frames (ORFs) encoding proteins that are similar in sequence to retroviral reverse transcriptases (Temin and Mizutami, 1970; Boeke and Corces, 1989) These genetic elements fall into several groups: (i) hepadnavirus of animals and caulimoviruses of plants, both DNA viruses; (ii) transposable elements which like retroviruses contain gag and pol genes and long terminal repeats (LTRs); (iii) certain fungal group II mitochondrial introns and a mitochondrial plasmid; and (iv) a group of transposable elements that also contain retroviral like gag and pol genes but do not contain LTRs (Xiong and Eickbush, 1990). Similarity in amino acid sequences of the reverse transcriptases of these elements suggests a common origin for many diverse reverse transcriptase sequences.

Although sequence similarities could be detected in other coding regions, the reverse transcriptase region is the only one common to all the elements and thus can be used for a comprehensive phylogenetic analysis of retroelements. Based on the seven peptide regions (domain 1-7) spanning 178 amino acids which are found to be common to all retroelements, the relationship between various retroelements has been established and a phylogenetic tree has been constructed (Xiong and Eickbush, 1990) as shown in Fig. 1.

The RNA-dependent RNA polymerases are all located on one branch which
Figure 1. The unrooted phylogenetic tree of the reverse transcriptase and RNA polymerase sequences (Xiong and Eickbush, 1990).
Elements of the same class located on the same branch of the tree are indicated by a box. RNA dependent RNA polymerase sequences are shown in shades of blue inside a quadrant separated by dots. Non-LTR retrotransposons are shown in shades of red and LTR-retrotransposons are shown in shades of green.
joins the reverse transcriptase (RT) branch on the segment connecting the non-LTR retrotransposons with the hepadnaviruses. The coliphages MS2 and SP are most distant members of this branch. The eukaryotic viruses clearly fall into the polio-like and the Sindbis-like groups. When these viral RNA polymerase sequences are used to root the tree, all RT containing elements fall into two major branches. One branch contains the bacterial msDNAs, group II introns and non-LTR retrotransposons while the second branch contains the three types of viruses (hepadnaviruses, caulimoviruses and retroviruses) and the LTR-containing retrotransposons. The overall organization of major types of retroelements is shown in Fig. 2. LTR retrotransposons contain, as the name implies, long terminal direct repeats of generally 200-500 nucleotides that flank a region of from 4.5 kb to 9.0 kb (Bingham and Zachar, 1989). The middle sequences contain one or more ORFs encoding the proteins necessary for replication and transposition of the element. The most complete elements contain three ORFs, called gag (group associated antigen), pol (polymerase) and env (envelope) after similar retroviral ORFs. Both the gag and pol ORFs encode polyproteins that are later cleaved into lower-molecular-weight proteins. The polyprotein encoded by gag is cleaved into smaller matrix, capsid and nucleocapsid proteins responsible for the protein-nucleic acid structure necessary for reverse transcription. The polyprotein encoded by pol is processed into: (a) reverse transcriptase, which copies the RNA copy of the retrotransposon into double stranded DNA; (b) aspartic protease, which cleaves the polyprotein into their component proteins (sometimes encoded by the gag gene); (c) RNaseH, which degrades the RNA template before second strand synthesis; and (d) integrase, which is responsible for inserting the double stranded DNA copy of the element into the host genome. The third ORF, env, encodes a membrane spanning protein that is intimately involved with infectivity in retroviruses. Although homologous env ORFs are apparently not present in retrotransposons, env-like ORFs are sometimes present and may encode analogous membrane spanning proteins, leading to the speculation that some retrotransposons may be infectious.
Figure 2. Major types of retroelements
Overall organisation of (A) a retrovirus, the avian leukosis virus, (B) a copia like retrotransposon, the yeast Ty1 element, (C) a gypsy like retrotransposon, the yeast Ty3 element and (D) a non-LTR retrotransposon, the Drosophila I factor. Open reading frames are depicted by blue and violet boxes. gag, encodes structural protein for the virion core; env, encodes a structural envelope protein; prot, protease involved in cleavage of primary translation products; RT, reverse transcriptase; RNase H, ribonuclease; endo, endonuclease necessary for integration in the host genome; LTR, long terminal repeats; PBS, primer binding site complementary to the 3' end of a host tRNA, used for the synthesis of first (-) DNA strand; PPT, polypurine tract used for the synthesis of the second (+) DNA strand. Green arrows show direct repeats of the host DNA created upon insertion. (Figure modified from Grandbastien, 1992).
The life cycle of LTR retrotransposons is similar in many regards to that of the retroviruses (Fig. 3). An integrated genomic copy of the element is transcribed by one of the cellular RNA polymerases, creating at once a mRNA encoding the proteins necessary for transposition and the template to synthesize a new DNA copy of the element. The mRNA is primed for reverse transcription by a specific cellular tRNA, and a double stranded DNA copy of the element is made by reverse transcriptase. Integrase inserts the new DNA copy into the host genome. The integration of the element is always flanked by a small target site duplication. Unlike DNA mediated transposable elements that use a cut-and-paste mechanism to move, RNA-mediated elements are necessarily replicative because the original copy, from which the mRNA has been transcribed, is not mobilized in the transposition process. Integrated copies can be lost, however, through an unrelated mechanism involving recombination between the LTRs (Bingham and Zachar, 1989).

The other large group of retrotransposons are those lacking long terminal repeats, sometimes referred to as retroposons (McClure, 1991). These elements contain reverse transcriptase domain but lack aspartic protease and integrase regions which, some also lack the gag like gene. Retroposons have a range of insertion site preferences; apparently some elements can insert randomly throughout a genome whereas others are restricted in their insertion to a single nucleotide sequence of a single gene. Their mechanism of reverse transcription and integration is quite distinct from that of the LTR retrotransposons and is facilitated by an endonuclease (which is sometimes termed integrase but is not homologous to the LTR retrotransposon and the retroviral integrases). The details of this mechanism have been described for R2 elements in the silkworm moth, *Bombyx mori*, and involve nicking of the target DNA and use of the exposed 3’ hydroxyl to prime reverse transcription (Luan et al., 1993). Second strand cleavage of the target then occurs followed by second strand synthesis of the element, perhaps mediated by the host DNA repair enzyme system. This system of transposition may operate in all non-LTR retrotransposons as well as SINE-like elements, perhaps exploiting pre existing nicks in the DNA or existence of
Figure 3. The life cycle of yeast Ty1 retrotransposon

The intracellular replication and integration of the retrotransposon is shown. U3, R and U5 regions within the LTRs are symbolised by pink, black and hatched boxes, respectively. Host DNA is indicated by a double line. The RNA transcript is shown as a wavy line. Virus like particle (VLP) structures are indicated by green hexagons. $A_N$ indicates the poly A tail of the transcript. The full length DNA is shown with the proposed terminal base pairs of the element brought together, but a covalent circular form is not implied. Drawing is not to the scale. [Figure modified from Boeke and Sandmeyer, (1991)]
an endonuclease with far less target site specificity than that of R element enzymes (Finnegan, 1989).

The SINES comprise a class of RNA mediated elements distinct from LTR retrotransposons and non-LTR retrotransposons. Not only they are short (ranging from 75 to as much as 500 bp in length) but also generally lack open reading frames and do not code for specific enzymes ("transposases") which mediate the insertion process. So they rather employ cellular mechanisms for retroposition. Work on vertebrates (in which SINES can account for upto 5% of the genome, as in the case of the human Alu sequences) have uncovered some general characteristics. They are present in families consisting of more than 100,000 individual members that are all of about the same length (except for the A-rich region at the 3' end) and exhibit 70 to >98% sequence homology. A given family is often represented by a consensus sequence, determined by sequencing a number of family members and aligning them to find the most common nucleotide at each position.

The generic SINE sequence contains an internal RNA polymerase III promoter, an A-rich 3' end (on the strand corresponding to the transcript), and flanking direct repeats. The A rich 3' end is quite variable in length and exact sequence and constitute the region of more heterogeneity among members of any given family. The A rich 3' end regions vary from less than 8 to longer than 50 bp and are often mixed with base pairs containing bases other than As. In fact, simple sequence repeats of the form (XAy)n, where X represents any other base, are often found in this region. Other 3' end patterns are also found, including other simple repeating sequences. Some bovine and goat families lack either A richness or even a simple tandem repeat structure. The direct repeats that flank the SINES are not a part of repeated DNA family member itself, but derive by duplication of target sequences at the site of integration. These direct repeats vary in size from a few base pairs to >30 bp in length and are generally A rich. In addition, as many as one third of the members of some families are not flanked by obvious direct
PLANT RETROTRANSPOSONS:

The first characterization of mobile DNAs in plants were initiated by Barbara McClintock who studied the terminal inverted repeat (TIR) transposons that first termed "controlling elements" (McClintock, 1951). Those were typical transposons not involving reverse transcriptase in their life cycle. Retroelements were not discovered in plants (Shepherd et al., 1984) until after their discovery in Drosophila (e.g. copia and gypsy) and yeast (e.g. Ty1 and Ty3). Despite the slow beginning, a large number of different plant retroelements have now been uncovered either by analysing the insertions in or near various genes (Pelissier et al., 1995) or by studies of cloned repetitive DNAs (Table 1). Many of the elements identified are of Ty1/copia group of plant retrotransposons partly because of frequent use of Ty1/copia specific primers to amplify and clone Ty1/copia relatives from plants (Flavell et al., 1992). However, Ty3/gypsy class of retrotransposons are also well represented. Only a few examples of LINE and SINE elements have been identified in plants, but this may be due to small number of targeted searches made for those classes of plant retroelements (Manninen and Schulman, 1993).

**Plant retrotransposons are ubiquitous, ancient, diverse and present in high copy number:**

Studies, relying on either the detection of retrotransposons or the isolation of highly repeated sequences, have demonstrated that retrotransposon derived sequences are an abundant component of many plant genomes (Voytas et al., 1992a) have shown the ubiquity of copia-like retrotransposons by detecting the expected PCR product using genomic DNA of plant species including representatives from 9 of 10 plant divisions. Also, the retrotransposons isolated from individual plants are usually highly diverse at DNA sequence level, far more so than similar retrotransposons amplified from Drosophila and yeast (Flavell et al., 1992). The diversity of retrotransposon isolated from the same individual plants cannot be explained due to their large copy number because even the
TABLE 1
SOME WELL CHARACTERIZED PLANT RETROELEMENTS
(Bennetzen, 1996)

<table>
<thead>
<tr>
<th>Name</th>
<th>Species</th>
<th>Total (kb)</th>
<th>LTRs (bp)</th>
<th>DRs (bp)</th>
<th>Copy No.</th>
<th>Type</th>
<th>Homologies/DRs</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bsl</td>
<td>Zea mays</td>
<td>3.2</td>
<td>302</td>
<td>5</td>
<td>0-5</td>
<td>-</td>
<td>pro.int,PMPA;two overlap</td>
</tr>
<tr>
<td>Cinl</td>
<td>Zea mays</td>
<td>-</td>
<td>690</td>
<td>5</td>
<td>1000</td>
<td>solo LTR</td>
<td></td>
</tr>
<tr>
<td>Cinful-1</td>
<td>Zea mays</td>
<td>8.6</td>
<td>586</td>
<td>5</td>
<td>nd</td>
<td>Ty3/gypsy</td>
<td>gag.gag.pro,RT,RNaseH,int.</td>
</tr>
<tr>
<td>Zeon-1</td>
<td>Zea mays</td>
<td>7.3</td>
<td>649, 663</td>
<td>5</td>
<td>250-1250</td>
<td>-</td>
<td>gag.one long</td>
</tr>
<tr>
<td>Hopscooch</td>
<td>Zea mays</td>
<td>4.8</td>
<td>231</td>
<td>5</td>
<td>nd</td>
<td>Ty3/gypsy</td>
<td>gag.gag.pro,RT,RNaseH; one</td>
</tr>
<tr>
<td>Megellan</td>
<td>Zea mays</td>
<td>5.7</td>
<td>341</td>
<td>nd</td>
<td>Ty3/gypsy</td>
<td>int,RT,nd</td>
<td>gag.pro,nd</td>
</tr>
<tr>
<td>Opt-1</td>
<td>Zea mays</td>
<td>~8.7</td>
<td>1256</td>
<td>5</td>
<td>30,000</td>
<td>nd</td>
<td>gag.pro,nd</td>
</tr>
<tr>
<td>PREM-2</td>
<td>Zea mays</td>
<td>9.4</td>
<td>1307</td>
<td>5</td>
<td>30,000</td>
<td>Ty3/gypsy</td>
<td>gag.pro,RT,RNaseH; two</td>
</tr>
<tr>
<td>Grande1</td>
<td>Z.diploperennis</td>
<td>13.8</td>
<td>631, 626</td>
<td>5</td>
<td>1300-1700</td>
<td>Ty3/gypsy</td>
<td>gag.gag.pro,RT,RNaseH,int</td>
</tr>
<tr>
<td>Osser</td>
<td>Volvox carteri</td>
<td>4.9</td>
<td>197</td>
<td>5</td>
<td>80</td>
<td>Ty3/gypsy</td>
<td>gag.pro,RT,RNaseH,int</td>
</tr>
<tr>
<td>Wis-2</td>
<td>Triticum aestivum</td>
<td>8.6</td>
<td>1755</td>
<td>5</td>
<td>200</td>
<td>Ty3/gypsy</td>
<td>gag.gag.pro,RT,RNaseH,</td>
</tr>
<tr>
<td>Lev-1</td>
<td>Sorghum bicolor</td>
<td>15.5</td>
<td>4560</td>
<td>5</td>
<td>400</td>
<td>Ty3/gypsy</td>
<td>gag.gag.pro,RT,RNaseH,nt;</td>
</tr>
<tr>
<td>Tst1</td>
<td>Solanum tuberosum</td>
<td>5.1</td>
<td>283</td>
<td>5</td>
<td>1</td>
<td>Ty3/gypsy</td>
<td>gag.gag.pro,RT,RNaseH;one</td>
</tr>
<tr>
<td>PDR1</td>
<td>Pisum sativum</td>
<td>5</td>
<td>156</td>
<td>5</td>
<td>50</td>
<td>nd</td>
<td>gag.pro,RT,RNaseH,</td>
</tr>
<tr>
<td>PG17</td>
<td>Prisma radicata</td>
<td>5.9</td>
<td>333</td>
<td>5</td>
<td>10000</td>
<td>Ty3/gypsy</td>
<td>gag.gag.pro,RT,int;two overlap</td>
</tr>
<tr>
<td>Tov-1</td>
<td>Oryza sativa</td>
<td>3.2</td>
<td>115</td>
<td>5</td>
<td>&gt;30</td>
<td>nd</td>
<td>gag.gag.pro,RT,RT;</td>
</tr>
<tr>
<td>Tov-1</td>
<td>Oryza sativa</td>
<td>3.2</td>
<td>574</td>
<td>5</td>
<td>30-300</td>
<td>nd</td>
<td>gag.gag.pro,RT,RT;</td>
</tr>
<tr>
<td>Tov1</td>
<td>Nicotiana tabacum</td>
<td>5.3</td>
<td>610</td>
<td>5</td>
<td>&gt;100</td>
<td>Ty3/gypsy</td>
<td>gag.gag.pro,RT,RNaseH,one</td>
</tr>
<tr>
<td>del1</td>
<td>Lilium henryi</td>
<td>9.3</td>
<td>2406-2415</td>
<td>5</td>
<td>&gt;40000</td>
<td>Ty3/gypsy</td>
<td>gag.gag.pro,RT,RNaseH; no long ORF</td>
</tr>
<tr>
<td>BARE-1</td>
<td>Hordeum vulgare</td>
<td>12.1</td>
<td>1829</td>
<td>4</td>
<td>5000-50,000</td>
<td>Ty3/gypsy</td>
<td>gag.gag.pro,RT,RT;</td>
</tr>
<tr>
<td>BIST</td>
<td>Hordeum vulgare</td>
<td>6.5</td>
<td>452</td>
<td>5</td>
<td>&gt;10000</td>
<td>nd</td>
<td>gag.gag.pro,RT,RNaseH,nt;</td>
</tr>
<tr>
<td>Toc1</td>
<td>Chlamydomonas reinhardtii</td>
<td>3.7</td>
<td>?</td>
<td>0</td>
<td>2-30</td>
<td>odd</td>
<td>RT</td>
</tr>
<tr>
<td>Athila</td>
<td>Arabidopsis thaliana</td>
<td>10.5</td>
<td>1539</td>
<td>5</td>
<td>&gt;30</td>
<td>-</td>
<td>none; two long ORFs</td>
</tr>
<tr>
<td>Tal1</td>
<td>Arabidopsis thaliana</td>
<td>5.2</td>
<td>514</td>
<td>5</td>
<td>1-3</td>
<td>Ty3/gypsy</td>
<td>gag.gag.pro,RT,RNaseH,one</td>
</tr>
<tr>
<td>Cin4</td>
<td>Zea mays</td>
<td>1-6.8</td>
<td>-</td>
<td>3-16</td>
<td>50-100</td>
<td>LINE</td>
<td>gag.gag.pro,RT,two ORFs on largest member</td>
</tr>
<tr>
<td>p-SINE1</td>
<td>Oryza sativa</td>
<td>0.125</td>
<td>-</td>
<td>14</td>
<td>SINE</td>
<td>(RNA);none</td>
<td></td>
</tr>
<tr>
<td>TS</td>
<td>N.tabacum</td>
<td>0.111-0.258</td>
<td>-</td>
<td>3-19</td>
<td>50,000</td>
<td>SINE</td>
<td>tRNA34;none</td>
</tr>
<tr>
<td>del2</td>
<td>Lilium speciosum</td>
<td>4.45</td>
<td>-</td>
<td>4-13</td>
<td>250000</td>
<td>LINE</td>
<td>gag;RT;two</td>
</tr>
<tr>
<td>SLm</td>
<td>Brassica napus</td>
<td>0.170</td>
<td>-</td>
<td>9-35</td>
<td>500</td>
<td>SINE</td>
<td>tRNA;none</td>
</tr>
<tr>
<td>Tal1-1</td>
<td>Arabidopsis thaliana</td>
<td>6.1</td>
<td>-</td>
<td>15</td>
<td>1-6</td>
<td>LINE</td>
<td>gag,RT;two overlap</td>
</tr>
</tbody>
</table>

Abbreviations:
- LTR - Long terminal repeat
- DRs - Flanking target direct repeats
- ORFs - Open reading frames
- pro - protease
- int - integrase
- PMPA - Plasma membrane proton ATPase
- nd - not determined
- RT - reverse transcriptase
- env - envelope protein
- LINE - long interspersed nuclear element
- SINE - short interspersed nuclear element.
extremely small *Ta* family of *copia*-like retrotransposons of *Arabidopsis thaliana* are also highly diverse (Flavell *et al.*, 1992a). It is more likely that the sequence heterogeneity reflects an ancient association between plants and *copia*-like retrotransposons. White *et al.* (1994) have uncovered retrotransposons insertions in the 5' and 3' regions of >30 normal plant genes using *copia*-like elements as query sequence in computer based sequence similarity searches of the DNA sequence databases. Fixation of retrotransponson sequences near normal genes, coupled with their degenerate nature provides additional evidence for an ancient association between *copia*-like retrotransposons and plant genomes. It is now known that plant genes are separated by huge regions of intergenic regions comprised of repetitive sequences. Of a 280 kbp region containing the maize adh1 gene, 197 kb comprises at least 37 classes of middle and highly repetitive DNA. A significant fraction of these sequences are thought to be of retrotransposon origin (Avramova *et al.*, 1995).

**Copy number:**

Plant retrotransposons have highly variable copy numbers ranging from single copies of *Tal* in *Arabidopsis* and of *Tst1* in *Solanum tuberosum* to >40,000 copies of dell in *Lilium longiflorum* (Grandbastien, 1992). Many other retrotransposons have been identified as extremely abundant sequences in plants. These studies indicate that the very large plant genomes may be the result of the ability of retrotransposons to attain phenomenal copy number by amplification. Over 20,000 copies of *BARE1, BIS1* and *dell* (Bennetzen, 1996) occur per haploid genome and they account for significant portions of their resident nuclear genomes. The tremendous potential of amplification via retrotransposition is illustrated in members of genus *Lilium*. The 14 species have enormous genomes of 30-45 Mkb which may be the result of unrestrained retrotransposition. 13,000 copies of the 9.35 kb *gypsy*-like retrotransposon dell comprises of 120,000 kb or 0.4% of the genome of *Lilium henryi*. *L. longiflorum* has >40,000 copies of dell which accounts for at
least 1% of its 34 Mkb genome. The 4.45 kb LINE del2, has 250,000 copies in Lilium speciosum which is equivalent to 4% of its genome (Wessler et al., 1995).

**Origin and Organization:**

From sequence comparisons of conserved region of retroelements, Xiong and Eickbush (1990) have inferred that since retrotransposons are the only elements common to both the LTR and non-LTR branches of the retroelement family, their structure is the most likely progenitor of all current retroelements. It suggests that retrotransposons arose early in the history of eukaryotes which partly accounts for their current broad distribution. Hence all major classes of retroelements may have arrived in plants by vertical transmission from an ancestor that arose prior to the emergence of plants.

Gel blot hybridization analysis have indicated that the organization of retrotransposons in the plant genome is such that most copies of these elements do not occur in tandem arrays, although rare genomic clones sometimes carry two or more copies of the same element (Bennetzen, 1996). In situ hybridization studies have indicated some patterns in the organization of plant retrotransposons. The BIS1 element was found to be missing from or reduced in centromeres, telomeres and nucleolar organizer of barley (Moore et al., 1991), while the Grandel element of maize was observed to be under-represented at centromeres and some what clustered in the distal regions of some chromosome arms (Palmgren, 1994).

**Retrotransposon structure variation:**

In plants, the structure of retrotransposons varies tremendously. Solo LTRs, presumably derived from unequal recombination between the LTRs of a single element, have been observed for a few plant retrotransposon families. Multiple nucleotide substitutions and small (1-4 bp) insertion/deletions are frequently observed. Large internal rearrangements are also reported (Bennetzen, 1996). Many related retrotransposons appear to contain unrelated internal sequences, suggesting
that some of them have acquired DNA from other sources. (Palmgren, 1994). Even the LTRs of two related elements in a species show extensive sequence variation. Due to the existence of processes for acquiring extra internal sequences and known inaccuracy of reverse transcriptase, retrotransposons appear to be evolving at much faster rates than the rest of their host plant genomes.

**Mechanism used by plant retrotransposons for retrotransposition:**

Based on the observation that LTR retrotransposons use the same basic mechanism of retrotransposition in fungi and insects, it can be safely conjectured that plant elements of this class transpose in basically the same way. This involves transcription of the integrated element into a full length RNA which is inserted into a virus like particle, the component of which are encoded by the RNA. The RNA is reverse transcribed by reverse transcriptase into extrachromosomal DNA, which then gets integrated into the genome by a specialized transposase (called integrase). The steps of this process and the transposition intermediates have been very well characterized for yeast and insect elements. Although little is known about plant elements, it is likely that analogous plant elements use similar intermediates. The retrotransposition mechanism of non-LTR retrotransposons is much less clear in animal and plant systems.

**Integration specificities of Retrotransposons:**

The replication cycle of retroviruses and their retrotransposon relatives requires the insertion of a DNA copy of their RNA genome into the host DNA. Although no specific DNA sequence is required at the site of integration, the distribution of insertion is non-random. Such insertion specificity is associated with TIR transposable elements in all species investigated, including plants where they are found to insert preferentially into or near genes (Cresse *et al.*, 1995). The degree of target site selectivity varies among different retroelements (Boeke and Corces, 1989). Some can apparently integrate at any location in the host DNA, while others
integrate almost exclusively at a limited set of sites. In the case of animals and fungi, retroviruses and retrotransposons show a strong preference for particular insertion sites (Ji et al., 1993). The yeast Ty1 and Ty3 elements insert primarily near tRNA genes (Chalker and Sandmeyer, 1992), while TART of Drosophila specifically inserts at its chromosome termini. Gel blot and in situ hybridization analyses have indicated that many retrotransposons are dispersed throughout their host genome except in centromeric regions in the case of at least some of the elements (Bennetzen, 1996). However, Pelissier et al., (1995) have shown that despite the presence of numerous euchromatic copies, the Athila element of Arabidopsis is concentrated in or near heterochromatic regions and have provided strong evidence that most of the heterochromatic elements retrotransposed directly into 180 bp satellite clusters.

Several elements (PREM1, Grande1 and Cinfull) were initially identified as insertions into the LTRs of other elements and clones of other elements with multiple copies of related elements have also been seen. In the maize genome, clusters of retroelements are abundant, in which they appear to be highly methylated and presumably heterochromatic. However, density centrifugation studies suggest that Cin4, Tal and Tntl are like Ac and Mu1 in their preferential association with genes containing isochores (Bennetzen, 1996).

Until recently, plant retroelements have only been associated with nuclear genomes but now, Ty1/copia and Ty3/gypsy group of elements have been found in mitochondrial genome of Arabidopsis (Knoop et al., 1996). It is not clear whether these elements initially entered the mitochondrial genome via retrotransposition or with other nuclear DNA acquired by the organelle genome, but most or all of the elements present in mitochondria are fragmented or otherwise defective. By sequence criteria, about 5% of the Arabidopsis mitochondrial genome appears to have a retroelemental origin.
Expression:

In spite of the isolation of numerous retrotransposons from various plants, not much is known about the expression of plant retrotransposons. This is because either most of them have been found as pre-existing insertions or most of the cloned ones are defective copies. The normal state of affairs for most of the thousands of plant retrotransposons present in a given plant is virtually undetectable rates of transposition. They are rarely (if ever) active during normal plant development. In contrast, retrotranscripts can be quite abundant in yeast (Curcio et al., 1988) and Drosophila (Bingham and Zacher, 1989). Initial studies of cloned retroelements from plants did not detect any expression, even in one case where element transposition had been detected. Subsequent studies have detected low levels of transcription for many elements (Bennetzen, 1996) often producing transcripts that can only be detected by amplification techniques (Hu et al., 1995). Transcripts homologous to some elements are found at relatively high levels and/or in some tissues. However, element-dependent expression has only been convincingly demonstrated in a few cases (Bennetzen, 1996). Even in the cases where element specific transcriptional initiation is observed, the low level of RNA detected suggests that only a small subset of the element population is being expressed.

Retrotransposon activity is inducible:

Retrotransposons are not active during normal plant development but they are responsible for new spontaneous mutations. This can be explained by the fact that transcription of several plant elements is inducible.

Transcription of Tnt1 element is induced when tobacco cells are treated with fungal extracts containing cell wall hydrolases (to remove the cell wall and produce protoplasts prior to cell culture initiation) (Pouteau et al., 1991). A component of fungal extract called an elicitor, was found to be specifically responsible for induction. Tnt1 can also be induced by a broad spectrum of microbial and fungal
elicitors all of which are able to activate the plant defense response: the hypersensitive response (Pouteau et al., 1994).

**Tissue Culture:**

Hirochika (1993) has amplified three new families of retrotransposons (Tto1-Tto3) using PCR from cDNA prepared from the protoplasts of an established tobacco cell line and have isolated a 5.5 kb long retrotransposon called Tto1. He showed that transcription of Tto1 was active only in cultured cells and was enhanced by protoplast formation. The copy number of Tto1 increased 10 fold in established cell lines, indicating that these retrotransposons may also be activated by tissue culture. In normally propagated plants, these three retrotransposons are cryptic as no difference in copy number was observed between individuals of same cultivars or even between different cultivars. Using the same technique of RT-PCR and RNA from cultured cells, the first active retrotransposons from rice have been isolated. Transcripts of three elements Tos10, Tos17 and Tos19 were detected in cultured cells but not in normal cells.

**Somaclonal Variation and Retrotransposon Induction:**

In *in vitro* cultures of plant cells, extensive genetic and cytogenetic modifications are induced. Such genetic variability observed in plants regenerated from tissue culture has been termed somaclonal variation and has been reported in most plant species investigated. Although somaclonal variations have been studied extensively as a source of plant improvement, little is known about its molecular basis. One possible mechanism suggested for somaclonal variation is that the activation of dormant DNA elements. Transposition of Ac and Spm elements has been detected in a few plants derived from cultured embryos lacking all such activity (Peschke et al., 1987; Peschke and Phillips, 1991), but the mechanism of activation has not been determined. It has been demonstrated that plants regenerated from cultured tobacco or rice cells possess new retrotransposon insertions making it
possible that activation of retrotransposons is responsible for somaclonal variations. Most tissue culture induced mutations are stable like the mutation induced by retrotransposons and the ubiquity of somaclonal variation may be explained by that of retrotransposons. Of the new insertions, it was found that seven of eight insertions of Tos17 were in single copy sequences, and four of these were in the coding regions of transcribed genes (Wessler et al., 1995). So given the frequency of these insertions, it is likely that retrotransposon induced mutations are a major factor in causing somaclonal variations.

**Effects of Retroelements on Plant Genome:**

Retroelements affect the genome in a number of ways. Many retroelements are known to cause gene inactivation by inserting within them, while some retroelement insertion into the genes have no obvious effect (Bennetzen, 1996). Multiple copies of retroelements within the plant genome may provide sites of recombination systems resulting into duplications, deletions, inversions and translocations. Since the LTRs contain transcriptional regulatory sequences, they provide the substrates for the evolution of gene regulation in plants. In animals and fungi, the LTRs of the retrotransposon are known to alter the expression of adjacent genes. Interestingly, by computer assisted searches of the Genbank and EMBL databases using three plant retrotransposons as query sequences, White et al. (1994) have shown the presence of ancient, degenerate retrotransposon insertions in close proximity of 21 previously sequenced normal plant genes. They have provided lines of evidence suggesting that some of the retrotransposon like sequences identified may influence the expression of adjacent genes. The retrotransposon like sequences in maize polygalactouronase (PG) gene contains sequence motifs that are common among genes expressed during pollen development (Allen and Lonsdale, 1993). In addition, a 501 bp fragment containing a positive regulatory region of a tomato gene expressed during pollen development (LAT59) (Twell et al., 1991) is composed entirely of a retrotransposon-like sequence. Surprisingly, only plant gene sequences
were identified as having significant similarity to either plant (*Tnt1*, *Hospscotch*) or *Drosophila* (*copia*) retrotransposons, when it is known that *copia*-like elements are highly expressed in *Drosophila* and cause many spontaneous mutations (Bingham and Zachar, 1989). In contrast, plant *copia*-like retrotransposons are transcribed at low levels under normal growth conditions and have been found to be responsible for only a few mutations. Because of the heterogeneous and ubiquitous nature of *copia*-like retrotransposons in plants, it is inferred that these elements may have an ancient association with plant genomes and may have had a longer time frame for insertion into the flanking regions of the genes to occur and become fixed. In contrast, analysis of both codon usage of *copia* and its phylogenetic relationship to other retrotransposons has led to hypothesis that *copia* like elements were horizontally transmitted to *Drosophila* or one of its ancestors, giving them shorter time to become fixed in the flanking region of the insect genes than in their plant counterparts. Retroelements that are not directly involved in regulation of gene expression by their not being present within or near genes, can still play a role in plant genome function and evolution. Each retrotransposon capable of retrotransposition is a potential source of a DNA sequence that can enhance the general mutation rate and in the situation that constitute what McClintock called 'genomic shocks', the activation of dormant retrotransposons could result in traumatic resetting of the genome (McClintock, 1984) which would provide highly altered progeny for natural selection to work on. This scenario is very important for the plants as they can not move to escape the unanticipated challenges of nature.

Retrotransposons make up such a large proportion of some plant genomes that it is natural to assume that they have major effects on them. Even if they do not disturb the genes directly, their mere presence can promote deletions and inversions of the genome by inter-element recombination. Dispersed elements necessarily promote duplication and loss of DNA in which they are embedded. Such phenomena are well documented in fungal and animal systems. It is yet not clear whether retrotransposons are a direct stimulus for the change in genome size or they simply
respond opportunistically to selection pressure favouring large genomes (Flavell, 1994).

Irrespective of the scenario, the fact is that retrotransposons are ubiquitous and have been surviving in plant genomes for a long evolutionary time span. Plants have evolved ways to deal with retrotransposons in such a way as to minimize their negative effects while keeping their positive role. Sequestration of most retrotransposons into methylated heterochromatic domains (Bennetzen et al., 1994) is one of the ways to minimize the negative impact of retrotransposons. Wessler et al., (1995) has proposed two models that can explain the retrotransposon activity with respect to plant genome function. One model envisages that retrotransposons are not transcribed during normal development because they are epigenetically silenced, perhaps by a mechanism analogous to the reversible inactivation of DNA elements or transgenes (Flavell, 1994a). It remains to be seen whether active retrotransposons are silenced by similar processes during plant development. The second model predicts that either the LTRs of plant retrotransposons contain transcriptional silencers or they lack cis-elements that can be recognised by trans-acting factors during normal development. Since it is known that the LTRs of Tnt1 and Tio1 contain cis-elements necessary for the induction of transcription in cultured or stressed cells (Grandbastien et al., 1994) this model seems plausible. Also, constructs containing the LTR of Tnt1 do not promote transcription if they are not induced. It is difficult to understand how all the LTRs have evolved as to go unrecognised during normal plant development, since there is a great diversity in LTRs of active elements. On the other hand retrotransposon expression during normal development may drastically affect the plant genome to such an extent that those elements are quickly eliminated from the population. An interesting case is presented by Arabidopsis that harbors quite a number of different retrotransposons (Konieczny et al., 1991) but their individual copy numbers are all low.
Thus *Arabidopsis* has managed to minimize the copy number of various retrotransposons it has been exposed to, thereby maintaining a small genome and also keeping diverse retrotransposons.

**Transcriptional Control of Retrotransposon Induction:**

Tobacco and rice elements are induced by cell culture, as expression in cell culture was the basis for their isolation. These elements are induced by other factors also. *Tnt1*, for example, is induced by microbial and fungal elicitors. Pouteau *et al.*, (1994) have shown that transcription of *Tnt1* can be activated by several microbial factors having a common ability to elicit a hypersensitive response in tobacco. These elicitors include Onozuka crude extracts from the fungus *Trichoderma viridae*, elicitors purified from *Phytophthora* fungal species and culture supernatants of the bacterium *Erwinia chrysanthemum*. They have also shown that the hypersensitive reaction prompted by tobacco mosaic virus, fail to promote *Tnt1* transcription induction. However it is possible that its induction is strictly localized to a few cells only. *Tnt1* expression is low in adult roots and almost absent in leaves, but is strongly induced in leaf-derived protoplasts (Pouteau *et al.*, 1991). This induction has been shown to be associated with the presence of cell wall hydrolases. *Tnt1* transcription starts in the 5'LTR and a major transcriptional start site has been determined (Pouteau *et al.*, 1991). Also, transgenic plants transformed with translational fusions between 5'LTR and β-glucuronidase reporter gene (LTR-GUS) (Jefferson *et al.*, 1987) showed a pattern of GUS activity similar to the pattern found for *Tnt1* expression, demonstrating that LTR sequences are sufficient to mediate protoplast specific expression. Casacuberta and Grandbastien, (1993) have localised the LTR sequences involved in the control of Tnt1 expression during protoplast preparation. Their result derived from transient expression analysis of LTR-GUS reveal the total loss of GUS activity with the construct in which most of the U3 region of the LTR had been deleted; although the construct contained both LTR transcriptional start sites and putative TATA boxes. A detailed analysis of the U3
region of the LTR reveals two particularly interesting sequences. In the distal region a 20 bp sequence called BI is present. BI displays dyad symmetry, contains in the centre a sequence (CATGTG) similar to the myc binding site (CACGTG) and the consensus sequence (CANNTG) for factor binding elements (E boxes) within the immunoglobulin enhancers. The sequence found (GGCATGTGC) is also very similar to a sequence in the Bz1 gene promoter from maize (GGCAGGTGC) that is essential for expression of Bz1 in the presence of maize regulatory genes, C1 and R. The core sequence also resembles the G box related sequences found in plant genes regulated by many different stimuli such as light or abscisic acid (Casacuberta et al., 1993). Another sequence, which is of 31 bp and called BII is found repeated, once in the distal and twice in the proximal region. BII does not show any clear similarity with any described transcription factor recognition boxes. By band shift analysis of BII box sequences, it has been shown that this BII box sequence interacts with tobacco nuclear proteins from protoplast but not from leaves. The proteins that interact with the box BII are thus induced during protoplast preparation. This lends credence to the hypothesis that this sequence plays a key role in the activation of Tntl transcription (Casacuberta et al., 1993).

In contrast to Tntl, Tto1 and Tto5 have been shown to get activated when tobacco is infected with tobacco mosaic virus (Wessler et al., 1995). A 38 bp repeat in the 574 bp 5'LTR of Tto1 is necessary for its induction following viral infection. The 38 bp repeat contains 13 bp repeat sequence which is involved in the activation of Tto1 during cell culture (Wessler et al., 1995). Finally, the rice Tos17 element, unlike other elements tested, remains induced during prolonged cell culture, and there is no evidence that Tos17 is activated by hypersensitive response.

**Are some retrotransposons actually retroviruses?**

The ubiquitous nature of retrotransposons in eukaryotes and the DNA sequence similarity along with the usage of the same proteins for the proliferation raises a question whether there are some forms of retrotransposons that are
equivalent to the integrating retroviruses i.e. one of these forms of retrotransposons that can cross the cellular barriers like retroviruses can do. For several LTR retrotransposons, intracellular virus like particles (VLP) have been observed, but they lack the envelope functions necessary for intercellular transmission. Till now, with one exception, retrotransposons have not been shown to be infectious agents like retroviruses. The exception is the discovery by Kim et al., (1994) that gypsy will infect Drosophila raised on a medium containing homogenised pupae from a gypsy-active Drosophila line. Sequence comparisons of different retroelements puts gypsy nearest to the retroviruses and so these are considered to be the best candidates for possible infectious activity.

It is known that retroviruses can acquire and transmit portions of their host genomes (transduction). Viruses take up prot oncogenes into viral genomes and convert them into oncogenes but they can take up other cellular functions also (Kuff et al., 1986). Acquired cellular genes have usually replaced an essential viral function thus inducing a functional virus into a defective one that requires a helper virus for infectivity. Proposed mechanism of host gene acquisition suggest that retroelements other than retroviruses have ability to acquire the host genes, but this has not been observed for any animal or fungal retrotransposon. In the retroelement world, it is not clear whether the acquisition of nuclear gene sequences is a unique feature of retroviruses. If it is, then elements such as Bs1 should be defective versions of still undetected plant retroviruses. The sequence of Bs1 shows it to be a defective element because part of the element's reverse transcriptase gene has apparently been replaced with a fragment of a cellular gene. Analysis indicated that Bs1 has acquired transmembrane domains of a maize proton ATPase gene, Mha1 and has selected for the conservation of the reading frame and amino-acid sequences of this acquired segment (Jin and Bennetzen, 1994). Also Bs1 belongs to Ty3/gypsy group of retrotransposons which are most like the retroviruses and the only group of retrotransposons to have demonstrated infectivity (Kim et al., 1994). The plasma membrane proton ATPase fragment within Bs1 has primarily undergone
Review of Literature

conservative mutations since acquisition from Mha1, suggesting a selection for function (Jin and Bennetzen, 1994). The portion of Mha1 acquired could specify attachment to plasma membrane, perhaps supplying an env function. Bennetzen (1996) has sequenced Cinful-1, an element of gypsy group which is the first plant retroelement to contain coding potential for all intracellular functions plus properly positioned additional sequences that could specify env proteins. Zeon-1 appears to be a defective copy of Cinful-1 that has replaced all but gag-coding region with additional sequences (Hu et al., 1995).

Other plant LTR retrotransposons have been found to contain long stretches of internal sequence with no apparent sequence similarity to known retroelement gene. By analogy with Bsl, studies are now underway to investigate whether these internal sequences might originate from normal cellular genes (Pelissier et al., 1995).

Finally, only the demonstration of infectivity of a retrotransposon can prove that it can act as a retrovirus, the potential infectivity of the best candidates of Ty3/gypsy group might have been overlooked due to their low copy number and low level of expression. So, searches for best candidates would be useful. In addition, an active retrovirus like element could be engineered from the components of retrotransposons and retroviruses. Such an element would provide excellent opportunities to study the mechanisms by which plants have managed to avoid or overcome retroviral pathogenesis which is so detrimental to animal kingdom.

**Evolution of Plant Retrotransposons:**

**Sequence Variation:**

Retrotransposons mutate at a high rate due to the propensity of reverse transcriptase to make base substitutions through copying error, and the possibility that an element does not need to code for active gene product. Its mutation rate has been estimated to be 1 in 7,000 to 1 in 50,000 residues per replication (Katz and Skalka, 1990). Thus even a few cycles of insertion, transcription, reverse transcription and reinsertion would lead to rapid divergence in the original sequence
of a retrotransposon. Its sequence could decay without selective constraint. Any retrotransposon transcript could be copied and reinserted as long as at least some active copies of reverse transcriptase and gag product (along with integrase for LTR elements) are present in a cell. At least, in the case of Bs1 of maize, it seems likely that these activities do not even need to arise from a homologous Bs1 element. All copies of Bs1 present in cells before and after a defined retrotansposition event were incapable of providing active gene products (Jin and Bennetzen, 1989).

Retrotransposons can also mutate as genomic inserts but the rate is much lower. The divergence of the LTR sequences of a particular retrotransposon reflects the time of insertion of that retrotransposon, because the two LTRs will always start out identical if during reverse transcription, the second template jump is always intramolecular (Hu and Temin, 1990). So the percent of divergence will indicate how long the element has been a resident within the genome. Consistent with this, the newly inserted Tnt1 and Bs1 elements have identical LTRs, while the two elements identified through their abundance, del1 and IFG, are apparently relatively ancient inserts.

Rearrangement:

In contrast to transposons, a retrotransposon gets fixed at its insertion site. However, LTR retrotransposons can rearrange to become solo LTRs by intramolecular recombination of two LTRs of one element, thus losing the internal sequence. This has been found at one site of insertion of Tal element in some races of Arabidopsis thaliana (Voytas et al., 1990). Also one solo LTR was found among a sample of twenty genomic clones of the dell element of Lilium henryi (Sentry and Smyth, 1989). Cin1 of maize represents the extreme case where it has been detected only as solo LTRs, as though it amplified a long time ago to have allowed all inserts to decay to this form (Shepherd et al., 1984).

Non-LTR retrotransposons are unlikely to be subject to such genomic rearrangements as they lack repeats. Their 5' deletions are likely to arise during reverse transcription and integration rather than as genomic inserts.
Dispersal and Horizontal Transmission:

Dispersal of plant retrotransposons envisages two aspects. First is the presence of the same element in various species. In most cases, plant retrotransposons are present not only in all the tested varieties of the same species but also in several related species, suggesting presence in a common ancestor before speciation events and subsequent vertical transmission. Some elements seem to be long term residents, such as Bsl and Wis-2 which are found in other related genera of Graminaceae family. Bsl of maize was recorded in all four species of genus Zea as well as in three related genera (Fuerstenberg and Johns, 1990). Similarly, Tnt1, a copia group element of tobacco was found in three other solanaceous species tested. Other elements such as dell, seems to have a fairly recent origin, as they are not found outside their genus. The dell was detected in 14 Lilium species examined but not in several other monocots. The number of copies of a particular element varies greatly between related host species. The dell, for example, is barely detectable in some Lilium species, while in others it occurs in thousands of copies (Joseph et al., 1990). The distribution pattern of a particular retrotransposon across species does not necessarily reflect their phylogeny. One possible interpretation for these patterns is that a retrotransposon may have infected a common ancestor of the lineage and subsequently undergone sporadic bursts of amplification in different branches. Another possibility, perhaps less likely, is that there have been repeated new 'infections' of members of the lineage over time. Also, both the possibilities are not mutually exclusive.

The second aspect is the distribution of different but related retrotransposons. Closely related elements are found in very divergent hosts. The plant elements Tal and Tnt1 are closely related (showing 52% amino acid sequence identity in their reverse transcriptase sequence) but their hosts, Arabidopsis and tobacco, respectively are in unrelated families. Similarly dell occurs in the monocot genus Lilium while its close gypsy group relative, IFG is abundant in the large genome of Pinus, a gymnosperm. Even, more strikingly so, elements of copia group closely related to
Tal and Tnt1 were, in fact, first discovered in Drosophila melanogaster (copia) and bakers' yeast (Ty1) and the same two host species carry representative gypsy group retrotransposons (gypsy, 412, 297 and 17.6 in Drosophila and Ty3 in yeast) as well. Also, the result of two phylogenetic studies on the presence of Ty1/copia type of reverse transcriptase sequences in plants is interesting. Ten different low copy number families of Tal-related retrotransposons have been characterized in Arabidopsis thaliana. These families fall into two groups, elements in the first (Tal to Ta7) being closely related to tobacco Tnt1 element, while those in the second group (Ta8 to Ta10) are most related to Drosophila retrotransposons (Konieczny et al., 1991). Therefore, within the same plant species, elements are found that one more closely related to insect elements than to other elements of the same species. This pattern can be explained by late, horizontal transmission for the second group (which would have evolved mostly outside the plant genome) from an element that was also ancestral to the Drosophila copia and 1731 elements. A similar analysis performed on a wide range of plant species has also demonstrated that many elements fall into subgroups that span species boundaries, so that the closest homologue of a given sequence is often from a different species (Grandbastien, 1992). Non LTR retrotransposons are also found scattered across a wide range of host species. Cin4 and del2 occur in monocot plants, LINES are abundant in mammals and the ribosomal inserts, R1 and R2 are present in insects. These findings imply that mechanisms other than vertical transmission also contribute to the spread of retroelements, and the horizontal transfer between species, and even across major taxonomic groups of organisms, is a widespread occurrence (Doolittle et al., 1989). The movement of retrotransposons between cells of different species require the presence of a vector. The most obvious candidates for vector are viruses. If retrotransposon transcripts were packaged in a viral capsid, they could be transferred passively to the other cells, individuals and perhaps to different species. If reverse transcriptase molecules were included with the transcript (as in retroviruses), a DNA copy could be made immediately rather than relying on the chance presence of
heterologous reverse transcriptase. Despite a very limited host range of many of the viruses, they may be able to enter the cells of a wide range of species, although they cannot undergo subsequent replication. In this case, the passenger retrotransposons can be transported and then be independent to amplify, if possible.

Whether horizontal transmission is taking place now, has not been tested so far, but it has occurred in the past. Ancestral horizontal transmission of retrotransposons does not imply that these elements represent degenerate retroviruses that have lost the ability to be infectious. Rather phylogenetic studies have demonstrated that both retroviruses and retrotransposons have, in fact, evolved independently from non-LTR retroposons (Xiong and Eickbush, 1990), thus supporting the Temin's proposition that retroelements have evolved from cellular DNA (Temin, 1980).

In contrast to various infectious retroviruses of animals, in plants their equivalent has not yet been found. The only examples of plant infectious elements using reverse transcription in their replication cycle are pararetroviruses, such as caulimoviruses which include cauliflower mosaic virus (CaMV), Caunmelina yellow mottle virus and rice tungro bacilliform virus. Despite their homology with the retroviral domains, caulimoviruses are not plant equivalent of animal retroviruses because they are DNA viruses, do not integrate into host genome, are closely related to Ty3/gypsy group of retrotransposon and lack env gene. Caulimoviruses and possibly other plant pararetroviruses, probably represent the acquisition of ancestral reverse transcription functions by pre existing viruses (Grandbastien, 1992).

**FUNGAL RETROTRANSPOSONS:**

**Yeast Retrotransposons:**

**Retrotransposons of *Saccharomyces cerevisiae***:

Five distinct families of transposable elements designated *Ty1-Ty5* have been identified in *Saccharomyces cerevisiae*. All the five families of elements are of retrotransposon types. Similar to retroviruses, they are flanked by long terminal
repeats (LTRs) and replicate by reverse transcription of an RNA intermediate. The LTR retrotransposons on the basis of internal sequence arrangement could be categorized into two distinct groups: (i) Ty1/copia and (ii) Ty3/gypsy group of retrotransposons after representative elements in *Saccharomyces cerevisiae*. In *Ty1/copia* group there are four different families *Ty1, Ty2, Ty4* and *Ty5*. *Ty1* and *Ty2* are closely related and share extensive sequence similarity. *Ty4* is a family of low copy number elements that are only distantly related to *Ty1* and *Ty2* and the one known member of *Ty5* family is an element found inserted near the left telomere of chromosome III (Voytas and Boeke, 1993). They are similar in overall structure. Each consists of an internal domain of several kb flanked by two LTRs of several hundred base pairs. The LTRs are analogous to LTRs of retroviruses and other retrotransposons. They contain promoter and 3'-end formation signals for transcription of genomic RNA. The internal domain of *Ty1* and *Ty2* contain additional sequences required for transcriptional activation, and the internal domains of all of the *Ty* elements encode proteins required for replication and integration. Replication occurs through reverse transcription primed from specific locations within the almost full length *Ty* transcript and results in regeneration of a full length DNA copy of the element (Fig. 3). Genomic insertions of *Ty* elements are flanked by 5-bp repeats of the host sequence upon integration. In yeast genome, isolated LTR sequences which are also flanked by 5-bp target repeats, are also present that can arise from intact elements by recombination. The solo LTR sequences of the *Ty* elements are more numerous than those of the complete elements. The LTR sequences of *Ty1* and *Ty2* have be designated as δ, of *Ty3* as γ and *Ty4* as τ. The sequences within the different *Ty* families are not equally conserved. *Ty1*, the largest family, seems to exhibit the highest degree of variation. Though *Ty2* differs from *Ty1* by two large regions, the predicted amino acid sequences of these two elements are fairly similar (Kingsman et al., 1981). *Ty1* and *Ty2* classes vary widely in sequence, and isolated elements with up to 30% divergence have been reported (Rothstein et al., 1987).
Transposition of both Ty1/copia and Ty3/gypsy group retrotransposons occurs by a mechanism that is strictly analogous to retroviral replication (Boeke, 1989). Ty element transposition has been studied in detail using plasmids that direct the overproduction of proteins and RNA from specific copies of Ty elements (Boeke and Corces, 1989). The plasmids bear fusion of the yeast GAL1 promoter to either Ty1 or Ty2 element, marked genetically to distinguish them from the other copies of Ty in the cell. Overproduction of these elements results in accumulation of virus-like particles (VLPs), which can be purified and used in biochemical analysis of transposition intermediates (Voytas and Boeke, 1993). These constructs, combined with sensitive detection systems for transposition events, have allowed detailed study of Ty transposition process. The life cycle begins with transcription of an element resident in the host genome. The resultant mRNA is translated to yield protein products necessary for replication and also serves as a template for the synthesis of new elements by reverse transcription. The two major polyproteins encoded by retrotransposons are equivalent to the products of gag and pol genes of retroviruses. The product of the gag gene assemble into a virus-like core particle in the cytoplasm. Packed inside this particle are the pol gene products, template mRNA and a specific tRNA that primes the DNA synthesis by reverse transcription. Reverse transcription generates a DNA copy of the element and this is integrated into the nuclear genome by integrase.

**Transcription:**

LTR retrotransposons are usually transcribed into a major RNA that extends from the transcription initiation signal located in the 5' LTR. Since both LTRs are typically identical, the production of terminally redundant transcripts that could serve as a substrate for transposition requires the suppression of termination signals located in the 5' LTR and of promoter components located in 3' repeat. Transcription signals for these retrotransposons are usually located in 5' LTR and inside the transcribed region. The Ty1 and Ty2 elements encode a major 5.7 kb RNA and two minor species of 5.0 and 2.0 kb. The 2.0 kb transcript is apparently 5' coterminal
with the 5.7 kb species. The status of 5.0 kb transcript is unclear, but in spt3 mutants, this transcript is clearly not coterminus (Boeke and Corces, 1989). Sequence analysis of a Ty2 element reveals the presence of two putative TATA homologies within the LTR, but only one of them, TATAAAA, located 74 bp upstream of the transcription start site, has an effect on the levels of Ty RNA. Deletion of this sequence reduces expression of the wild type Ty RNA five to seven folds without changing the 5'end of the transcript (Liao et al., 1987). The rate of transcription initiation of this RNA is regulated by an upstream activating sequence (UAS), also located within the LTR, 111-140 bp upstream to the transcription initiation site. This element, unlike UASs described in other yeast genes, is necessary but not sufficient to promote transcription when positioned upstream of the TATA box; deletion of part or all of the UAS greatly reduces transcription without affecting the mRNA start site. Two additional regulatory sequences are located in the internal epsilon region of the Ty2 element. These sites, unlike the UASs, are similar to eukaryotic enhancers in that they can stimulate transcription from a site within the transcribed region (Liao et al., 1987). Similar dissection of Ty1 element showed an essential TATA element but no evidence for a UAS in the LTR. Instead, sequences just downstream of the LTR that were required for high Ty RNA level were identified (Fulton et al., 1988).

Regions of homology to core and SV40 enhancer sequences within Ty elements have been reported (Errede et al., 1987; Roeder et al., 1985). These sequences may be responsible for the effect of Ty elements on adjacent gene transcription; they may also be involved in the transcriptional control of Ty elements themselves. Competition between Ty promoters and adjacent promoters may be an important factor.

Host Factors:

Ty elements often transpose into the 5' regulatory sequences of genes, causing mutant phenotypes by inactivating or stimulating the expression of the genes. In mutation that result in over expression of the adjacent gene, transcription of the Ty is divergent from that of the overproduced gene, and the expression of both is
controlled by genes at the mating type (MAT) locus. In haploid cells or in a/a or a/a diploids, the gene adjacent to Ty is expressed at levels 5-20 fold higher than in the MATa/a diploids (Errede et al., 1980) This suggests that Ty regulates transcription from an upstream position in a manner resembling the effect of enhancer sequences.

In addition to MAT, expression of Ty elements and their effect on adjacent gene expression depends on several other host genes. Some of these regulators, such as STE7, STE11 and STE12 genes, were identified via sterilizing mutations that prevent mating in a and a cells; they may encode proteins involved in Ty transcription. Two proteins that bind specific sequences internal to Ty elements have been identified (Company et al., 1988; Goel and Pealuman, 1988). These proteins are probably important in the transcriptional control of this retrotransposon.

Additional host genes that are involved in the expression of the Ty element have been identified by the ability of mutation in these host genes to reverse the phenotype of Ty induced mutations. About 15 SPT (suppressor of Ty) genes have been identified that suppress the phenotype due to insertion of solo LTR, complete Ty element or both (Boeke and Corces, 1989).

**Target Site specificities of Yeast retrotransposons:**

Yeast retrotransposons show target site specificity. Among all, the Ty3 elements are the best examples. Integration sites of Ty3 are limited to the upstream regions of the genes transcribed by RNA polymerase III (pol III) (Voytas and Boeke, 1993). Using the plasmid target assay, it has been shown that in case of tRNA, 5S and U6 genes, which are transcribed by pol III, Ty3 integrates within 1-4 nucleotides of start site of transcription. This target preference is dependent on functional promoter element within the gene transcribed by pol III. Altering the start site of pol III transcription changes the site specificity of integration, suggesting that Ty3 element interacts with some component(s) of the pol III transcription machinery probably TFIIB (Chalker and Sandmeyer, 1993).
The insertion specificity of Ty1 has been studied using an entire chromosome III of yeast, *Saccharomyces cerevisiae* as a target. Ty1 transposition was shown to be located on a few discreet regions of the chromosome and surprisingly, most insertions were clustered upstream of tRNA genes and more than half (57%) were within 400 bp of a tRNA gene. Also, *in vitro* transposition assays indicate that this target specificity was not dictated by nucleotide sequence. The association of Ty1 integration sites with upstream regions of tRNA genes suggests interaction between the Ty1 integration apparatus and either the pol III transcriptional machinery or the chromatin upstream of pol III transcribed genes.

The Ty1 elements have also been found to be present as pre-existing insertions on chromosome III reflecting their affinity for the tRNA genes. The native chromosome contains multiple Ty1, Ty2, Ty4 and Ty5 insertions and most of these, with the exception of Ty5 element are closely associated with tRNA genes. Further characterization of the mechanisms that dictate the specificity of integration may reveal a new set of host factors that interact with retrotransposons.

**Retrotransposons of fission yeast, *Schizosaccharomyces pombe*:**

Two related families of retrotransposons have been isolated from *S. pombe* and named as *TfJ* (Transposon of fission yeast1) and *TfJ2* (Levin and Boeke, 1990). *TfJ* and *TfJ2* both possess long terminal repeats (LTRs) and sequences encoding amino acids homologous to protease, reverse transcriptase and integrase domains of retroviruses. The chromosomal locations and copy number of *TfJ* and *TfJ2* differs greatly in various isolates of *S. pombe*. The complete sequence of *TfJ* shows a unique aspect that it contains a single open reading frame whereas other retroviruses and retrotransposons usually possess two or more open reading frames. *TfJ* has been shown to be transpositionally active and producing VLPs (Levin and Boeke, 1993).
Review of Literature

**Retrotransposons of other fungi:**

Various retrotransposons or retrotransposon-like sequences have been identified in fungal species other than *Saccharomyces cerevisiae* and *Schizosaccharomyces pombe*. These include members of both LTR family and non-LTR family. The members have been isolated from various fungal genera. McHale *et al.*, (1992) isolated the first LTR retrotransposon, *Cfi-1* from the plant pathogenic fungus *Cladosporium fulvum*. They identified two direct terminal repeats, similar in size and sequence to retroviral LTRs, and predicted three long open reading frames (ORFs), two of which show homology to vertebrate retroviral *gag* and *pol* genes. Virus-like particles co-sedimenting with reverse transcriptase activity in homogenates of the fungus were observed, although this did not provide conclusive evidence for functional transposition of the *Cfi-1* element in the fungal genome.

*Magnaporthe grisea* is a filamentous ascomycete that parasitizes a wide variety of gramineous hosts. The genomes of *M. grisea* strains which are pathogenic on rice contain a family of dispersed, repeated DNA sequences named MGR (Hamer *et al.*, 1989; Valent and Chumley, 1991). The molecular analysis of these repeated sequences allowed the identification of a retroelement from *M. grisea* (Dobinson *et al.*, 1993). The element, designated *grasshopper* (*grh*), was present in multiple copies throughout the genome and contained two long terminal repeats of 198 bp plus one predicted ORF with sequences homologous to the reverse transcriptase, RNase H, and integrase domains of retroelement *pol* genes. Another retrotransposon from this species, named MAGGY, is 5638 bp long and has LTRs of 253 bp flanking the two ORFs (Leong *et al.*, 1994). Both *M. grisea* elements MAGGY and *grh* belong to the gypsy family. From fungal plant pathogen, *Fusarium oxysporum*, the retrotransposon, *skippy* has been isolated which is 7846 bp in length, is flanked by identical LTRs of 425 bp and target site duplication of 5 bp (Anaya and Roncero 1995). *Skippy* is closely related to elements of gypsy group. *Boty*, which belongs to gypsy group has been isolated from phytopathogenetic fungus, *Botrytis cinerea* (Diolez *et al.*, 1995). It is present in multiple copies and distributed throughout the
genome of some of its isolates. It has two direct repeats of 596 bp. A 9.4 kb LTR retrotransposon called Ylt1 has been isolated from dimorphic fungus, Yarrowia lipolytica. It is bounded by LTRs called Zeta elements which are of 714 bp and have signals for start and termination of transcription. Its copy number is dependant on the strain but at least 30 copies of complete element and 30 of zeta elements per haploid genome are present (Schmid-Berga et al., 1994).

The presence of copia-like elements has also been shown in plant pathogen Phytophthora infestans using degenerate primers in a PCR reaction (Tooley and Garfinkel, 1996). An LTR containing family called Fosbury from Pyricularia grisea (Schull and Hamer, 1996) resembles repeat sequence MGR586 and is present in isolates that infected rice. Fosbury and MGR586 are sometimes associated in the genome. Tcal of Candids albicans is a 5.5 kb retrotransposon and shows 5 bp target site duplication (Chen and Fonzi, 1992). Filamentous fungus Fusarium oxysporum contains Foret-l which belongs to gypsy class (Julien et al., 1992). CgTl, which is a non-LTR retrotransposon has been isolated from Colletotrichum gloeosporioides (He et al., 1996). Its length is 5.7 kb and its 30 copies are dispersed in the genome. Its termini have 3’ A rich domain and insertion site of one of its copy is flanked by 13 bp direct repeat. Tad of Neurospora also belongs to the non-LTR family (Kinsey and Helber 1989). It has been found in all strains of Neurospora but the active copy has been found in strain Adiopodoume from ivory coast (Sewell and Kinsey, 1996). It has two ORFs, ORF1 codes for the putative protein of 786 amino acids and ORF2 code for 1156 amino acids which has homology to reverse transcriptase (Cambareri et al., 1994). Prtl, a 4.7 kb long retrotransposon like sequence from filamentous fungus, Phycomyces blakesleeanus is another member of the non LTR family. Instead of LTRs it has short (54 bp) terminal inverted repeats. It is part as a single copy is blakesleeanus genome and shows no target site duplication (Ruiz-Perez et al., 1996)

In addition to their presence in nuclear genome of fungi, the reverse transcriptase-like sequences are found in mitochondria of Neurospora. The
Mauriceville and Varkud plasmids of *Neurospora* are closely related DNA elements whose nucleotide sequence and genetic organization suggest relationship to retrotransposons and mitochondrial introns. Both plasmids potentially encode a reverse transcriptase protein of 710 amino acids (Akins *et al.*, 1988).

**ARTHROPOD RETROTRANSPOSONS:**

**THE LTR-containing family**

The *Ty3/gypsy* family

The *Ty3/gypsy* family is a large LTR-containing family in arthropods and other organisms. The canonical *gypsy* element in *Drosophila melanogaster* is 7.4 kb long with 482-nucleotide LTRs and creates a 4 bp target site duplication (Marlor *et al.*, 1986; Yuki *et al.*, 1986). The corresponding yeast element is *Ty3* (Boeke, 1989). The gene structure of this family is similar to that of retroviruses in that the integrase domain follows the reverse transcriptase domain.

In *Drosophila melanogaster*, this family is represented by the elements *gypsy* or *mdg4*, 412, 17.6, 297 and *micropia*, among others. Members isolated from other *Drosophila* species include *gypsy* from *D. virilis*, *Ulysses* from *D. virilis* and *D. subobscura* and *tom* from *D. ananassae*. Others have been detected using Southern blot analysis in many other *Drosophila* species. Among Lepidoptera, *Bombyx mori* contains at least two family members, represented by the *Mag* and BM-A.L11 element, and the *TED* element has been isolated from *Trichoplusia ni*, as an insertion into the genome of the Autographa californica nuclear polyhedrosis virus cultured in *T. ni* cells. Another member of the gypsy family called *Woot*, has been isolated from red flour beetle, *Tribolium castaneum*. It is unusual in having extremely long LTRs (3.5 kb) but is otherwise similar to *Ulysses*. Given the diversity of these elements in *D. melanogaster* alone, other arthropods should harbor a wide variety of retrotransposons (Robertson and Lampe, 1995).

Recent work has indicated that *gypsy* may be a bona fide retrovirus. LTR retrotransposons have long been known to resemble retroviruses in structure, in the
degree of amino acid similarity of their various proteins, and in the way that many of
them form virus-like particles (VLPs) prior to integration into the genome (Bingham
and Zachar, 1989). They are, however, not classified as retroviruses because they
lack the crucial property of infectivity. Some Ty3/gypsy elements (e.g. gypsy, TED
and 17.6) have a third ORF that encodes a presumed membrane-spanning protein
resembling the retroviral env ORF, indicating that at least some may be able to leave
the confines of their host cells. Syomin et al., (1993) have found an extracellular
form of a gypsy VLP that contains full length polyARNA-DNA complex. Kim et al.,
(1994) transferred active gypsy elements from D. melanogaster stocks (SS strains) to
those without active elements (MSN strains) by injection and by feeding MSN larvae
a diet containing ground SS strain pupae. If these results survive scrutiny, gypsy
would become the first nonvertebrate retrovirus.

The Ty1/Copia Family

Ty1/copia is the other large group of LTR retrotransposons with an extremely
wide host range, including insects (Bingham and Zachar, 1989), yeast (Ty1)(Boeke,
1989), and many plants (Flavell et al., 1992b). The archetype of this family is the
Drosophila element copia, which was first described by Finnegan et al., (1978) from
a very abundant RNA species in tissue culture cells. The copia element is 5.15 kb in
length with 276-bp direct terminal repeats and produces a 5-bp target site duplication
of host DNA upon insertion (Dunsmuir et al., 1980; Mount and Rubin 1985). A
diagnostic feature of these elements is the presence of the integrase domain
preceding the reverse transcriptase domain, which differs from all other LTR
retrotransposons (Fig. 2). Copia elements have been isolated and sequenced from D.
melanogaster and D. simulans and Southern blot analysis have shown that they are
widespread in the genus. D.melanogaster contains one other copia-like
retrotransposon, 1731 element (Robertson and Lampe 1995).

The best distributional and phylogenetic analysis of copia-like elements
comes not from arthropods but from plants (Flavell et al., 1992b); however, these
results may be relevant to arthropods, once a greater number of species have been examined more closely. The use of PCR primers derived from the conserved regions of the reverse transcriptase domain reveals that diverse copia-like elements are essentially ubiquitous in plant genomes. Both vertical and horizontal modes of inheritance could be inferred from the sequences of the PCR products, although vertical inheritance appears to predominate. In some cases, however, species divergent by at least 500 million years contained copia elements with 65% amino acid identity, indicating that horizontal transfer of copia-like elements must take place, albeit infrequently (Flavell et al., 1992b).

The Pao Element

A third distinct type of LTR retrotransposon described from Bombyx mori belongs to neither the Ty1/copia family nor the Ty3/gypsy family. The Pao element was detected as an insertion into a ribosomal DNA (rDNA) spacer unit. It is 4.8 kb in length, contain 634-bp LTRs, and apparently produces a 5 bp target site duplication upon insertion. It contains a single ORF that includes gag and pol domains and codes for 1158 amino acids. Its reverse transcriptase sequence suggests that Pao is most related to the TAS element of the intestinal roundworm Ascaris lumbricoides (Xiong et al., 1993) and that it forms a third distinct branch of the LTR retrotransposon group.

The NON-LTR Family

The non-LTR retrotransposons are related to the long interspersed nuclear elements (LINEs) first described in vertebrates, where they can account for up to several percent of the genome. These elements contain relatively long unique 5' untranslated sequences that precede a central coding region containing one or two ORFs coding for gag-like and pol-like proteins. A second untranslated region at the 3' end containing a poly A tract of variable length is preceded by one or more copies of the polyadenylation signal sequence AATAAA (Hutchinson et al., 1989). These
Review of Literature

retrotransposons induce target site duplications ranging from 4 to 22 bp upon insertion. They are often present in the genome in various 5' truncated forms with few intact copies present, probably as a result of incomplete first strand synthesis or incomplete transcription of the element.

In D. melanogaster, the non-LTR elements include the I, F, G, R, jockey, Doc, HeT-A and TART elements (Robertson and Lampe, 1995). These elements can be as long as 13 kb (TART), but most are between 4 and 5 kb and are present in the genome in about 10 to 50 copies. Related non-LTR elements have also been characterized from other insects, including Juan-A from Aedes aegypti, Juan-C from Culex pipiens, T1 and Q from A. gambiae, NLR1Cth from C. thummi, and Dong from Bombyx mori. Other related elements have been detected via Southern blot hybridization in the bees, Apis mellifera and Apis cerana, and the moths, B. mori and Antheraea yamamai. The alignments of the reverse transcriptase ORFs of most of these elements, and their deep branches in phylogenetic trees, suggest that each represents a subfamily of elements. Although their phylogenetic relationships are incongruent with the host phylogenies, the depth of the branches indicates that any horizontal transfers occurred long ago (Robertson and Lampe, 1995). The exception is one Drosophila non-LTR element. Mizrokhi and Mazo (1990) detected jockey elements in just a single species outside the melanogaster species group, D. funebris. The jockey element could not be detected in its sibling species, even under low stringency conditions, indicating it had not been inherited vertically.

The R elements (R1 and R2) are an unusual group originally discovered as insertions into a subset of the 28S rRNA genes in D. melanogaster (Long and Dawid 1980). They have been found in at least nine insect orders representing both the Hemi- and Holometabola, and in 43 of 47 species examined, they interrupted and thereby destroyed between 5 and 40% of the 28S rRNA genes (Jakubczak et al., 1991). R elements have also been detected in the Diplopoda, Chilopoda, and Arachnida. Full-length R1 elements are 5.35 kb and contain two overlapping ORFs similar to retroviral gag and pol genes, respectively. R2 elements are 3.6 kb long and
contain one ORF that appears to be a \textit{gag-pol} fusion. The reverse transcriptase encoded by the gene is also associated with an endonuclease activity but lacks RnaseH activity (Luan \textit{et al.}, 1993). R elements are unusual because of their extraordinary degree of insertion-site specifically. All \textit{R2} elements insert between a particular pair of nucleotides in the \textit{D. melanogaster} 28S rRNA gene, while all \textit{R1} elements insert exactly 74 bp downstream (Jakubczak \textit{et al.}, 1990). In both \textit{B. mori} and \textit{D. melanogaster} copies of these elements are more than 98\% identical at the DNA level. Since they are present only in a subset of the rDNA genes, Jakubczak \textit{et al.}, (1991) have suggested that most copies are ultimately lost through unequal crossing over, with new copies inserted by transposition, which accounts for the high degree of identity among them. Sometimes, however, quite distinct \textit{R1} or \textit{R2} elements could be found in a single species (Burke \textit{et al.}, 1993). These cases are thought to represent the persistence of polymorphic forms of R elements in a genome rather than horizontal transfer, primarily because extremely similar R elements have not been found in phylogenetically distant species.

The Het-A and TART elements are unusual in that they perform a genuine cellular function. Most eukaryotes contain an enzyme system (telomerase) for the addition of short repeated sequences to the ends of chromosomes to preserve their integrity (Zakian, 1989). These repeats have a sequence similar to TTNGGG and have been detected via Southern blot analysis in a wide variety of eukaryotes including arthropods (Okazaki \textit{et al.}, 1993). Noticeable exceptions are the Diptera, some Coleoptera, and other sporadically distributed insect species. In \textit{D. melanogaster}, Het-A and TART transpose preferentially into the ends of chromosomes and apparently constitute their telomeres. Het-A elements are 6 kb and encode only a \textit{gag-like} protein in one ORF (Biessmann \textit{et al.}, 1992). TART (telomere associated retrotransposon) elements are longer than Het-A element, but their exact length is undetermined. Elements have been cloned that are >13 kb and that have \textit{gag} and reverse transcriptase domains. Both Het-A and TART have unusually long 3' untranslated regions. They also share several unique properties of
transposition. They transpose preferentially to telomere ends, always have their 5' ends at the distal end of the chromosome, and apparently do not produce a target-site duplication (Levis et al., 1993).

MAMMALIAN RETROTRANSPOSONS:

A variety of mammalian retrotransposons have been described. These include the families belonging to both LTR and non-LTR retrotransposons. These derive primarily from rodent and primate hosts. Because the host genomes are also inhabited by a variety of endogenous retroviruses, it is sometimes difficult to differentiate between endogenous retroviruses and retrotransposons.

Rodents:

Mouse: IAP (Intracisternal A-particle) sequences are endogenous retroviral-like retrotransposons, present at 1,000 copies in the mouse genome. A copy encodes nearly intact gag and pol genes. IAP expression is controlled at transcriptional level. With the help of transgenic mice the expression of an IAP has been shown to be restricted to male germ line (Dupressoir and Heidmann 1996). A family of 5.7 kb retrotransposons called ETn (Early Transposon) having two subfamilies is known. One element of a subfamily present in the immunoglobulin gamma 1 switch region has 5' and 3' LTRs and a 6 bp target site duplication. Southern hybridization suggests that both subfamilies of ETn sequences are represented many times in the mouse genome (Shell et al., 1990). Another abundant retrotransposon family inhabiting mouse genome is the VL30 group. A member of this family is 4834 bp with LTRs of 568 bp (Adams et al., 1988). No intact ORFs have been found although some region show homology to gag and pol ORF. Nevertheless, it has priming sites and viral packaging sites suggesting VL30 relies on protein functions of other retroelements. Peromyscus leucopus, the white footed mouse, contains mys family of retrotransposons having 500-1000 copies per haploid genome. The elements of mys family were not found in house mouse Mus musculus. The elements tend to insert
specifically just upstream of a well conserved 11 bp sequence (Pine et al., 1988).

**Rat:** A full length rat IAP, that has undergone transposition, has been identified and characterized. It is 6253 bp in length and has typical retroviral structures including *gag*, *pol* and *env* domains with flanking LTRs. Its 580 copies are randomly distributed in the rat genome (Xiao et al., 1995). From rat cardiac cDNA library, eleven cDNA, clones of retroviral origin have been isolated. One of them, SORO-1, is approximately 3.5 kb and contains a LTR. SORO-1 hybridizes to a mRNA of approximately 7 kb that is present both in rat heart and liver (Martin et al., 1995). Another family, nerve growth factor-inducible cAMP-extinguishable retrovirus-like (NICER) element family, has been identified. The family consists of several hundred members. d5 is one member which is induced by nerve growth factor (NGF) in rat adrenal pheochromocytoma PC-12 cells and this induction is repressed by cAMP (Cho et al., 1990).

**Hamster:** The retrotransposon-like elements of the intracisternal A-particle (IAP) sequences occur in about 900 copies per haploid hamster cell genome. By applying fluorescent *in situ* hybridization (FISH), these elements were found to be distributed in specific patterns in many of the 44 hamster chromosomes. In Syrian hamster cells cultured *in vitro*, retrovirus-like intracisternal R-type (IRP) particles are present, the genes of which are distinct from those of IAP (Altenschildesche et al., 1996).

**Human:**

One group of human retrotransposon constitute truncated HERV elements that contribute the majority of known human endogenous retroviral sequences. The HERV-H family, for instance, consist of at least 800-1000 truncated elements of about 5.8 kb in length, with large deletions in the *pol* gene and a completely deleted *env*. Only a small subpopulation of RTVL-H elements, about 5-10% contain full length *pol* and *env* sequence (Leib-Mosch and Seifarth, 1996). Another group of putative retrotransposons is the THE-1 sequences (transposon-like human elements),
which has been estimated to comprise approximately 10,000 copies in human genome. Together with additional 30,000 solitary LTRs these sequences make about 1% of the total genomic DNA. The consensus THE-1 element is 2.3 kb in length and is flanked by LTRs of 350 bp. However, these elements lack ORFs and display no detectable sequences homology with other retrotransposons or retroviruses. As evidence of THE-1 retrotransposition, extra chromosomal circular DNAs containing THE-1 sequences were detected in several cell lines. A family of non-LTR retrotransposons, called LINE-1 family (long interspersed elements), has been identified in human genome. It consists of approximately 100,000 copies. The majority of these elements are truncated at the 5' end. Full length LINE-1 elements are approximately 6.1 kb in size and are estimated to comprise about 3500 copies in the human genome. Characteristic features of LINE-1 are an internal polymerase II promoter in the 5' untranslated region and a polyadenosine tract at the 3' end, which suggests that LINE-1 elements are generated by reverse transcription of polyadenylated mRNAs. Another type of non-LTR retrotransposons is the short interspersed element (SINES). The most abundant and best studied SINE family in primates is the family of Alu repeats which make up about 5% of human genome. They are approximately 300 bp in length and generally have well-defined 5' and 3' ends. In addition, 4000-5000 copies of an element termed SINE-R are present in human genome which are derived from the LTR of the human endogenous retrovirus HERV-K (Leib-Mosch and Seifarth, 1996).

**AVIAN RETROTRANSPOSONS:**

In birds, a family of non-LTR retrotransposon has been identified in chicken genome (Haas et al., 1997). The family is composed of CR1 (chicken repeat) elements, 30000 copies of which are present scattered among single copy sequences in the genome. The 3' end of these elements conform to the consensus (C/GATTCTRT)\textsubscript{1-3} instead of usual poly A tracts. The conceptual translation product of ORF2 is predicted to contain endonuclease domain in addition to a reverse
transcription domain. Phylogenetic analysis of 52 CR1 elements with most complete 3' ends divides them into six subfamilies designated as A-F. The consensus sequence of subfamilies showed considerable divergence (Vandergon and Reitman, 1994), implying that the CR1 subfamilies are ancient. Another retrotransposon called ART-CH has been characterized from chicken genome. It has the structure of 3300 bp long provirus, including two 388 bp long terminal repeats (LTRs), a tRNA (Trp) binding site, and a polypurine tract. At least some of the 50 genomic copies of ART-CH are transcribed into polyadenylated RNA and several short region of homology with gag-related sequences. Based on its features ART-CH has been grouped with the VL30 retrotransposons of rodents (Vandergon and Reitman, 1994).

**AMPHIBIAN RETROTRANSPOSONS:**

Members of two related families of transposable elements, Tx1 and Tx2, have been isolated from the genome of *Xenopus laevis* (African clawed frog), and characterized. The smaller version in each family (Tx1d and Tx2d) show characteristics of DNA elements. The longer version Tx1c and Tx2c) have internal segment of 6.9 kb that includes two long ORFs. ORF1 shows homology with the gag and ORF2 shows homology with pol of non-LTR retrotransposons, suggesting that these elements are more closely related to non-LTR type elements (Garrett et al., 1989).

Another retrotransposon like element called IA11 has been isolated from *Xenopus laevis* genome as a cDNA clone. Its sequence of 4.5 has a 220 bp repeat at its ends indicative of a retrotransposon-like structure. Multiple copies of IA11-related sequences were found in the *Xenopus* genome, constituting solo LTRs of 1267 bp and unique region copies (i.e. sequences internal to the repeats in the cDNA). Inverted repeats of 5 bp and apparent target site duplications of 5 bp surround the sequenced solo LTR (Greene et al., 1993). Hsr1, a family of long repetitive DNA sequences have been studied in the genome of terrestrial *Salamander hydromantes*. The sequence analysis of a 5 kb fragment (Hsr1A) of one member has
identified it as a member of Ty3/gypsy class of retrotransposons. About $10^6$ Hsr1 sequences are present in the large hydromantes genome which is the highest number of copies so far discovered for retrotransposon-like elements in a eukaryote (Marracci et al., 1996). Tpa1 and Tpa2 are the sequences belonging to Ty1/copia group retrotransposons have been isolated from an amphibian Pyxicephalus adspersa. Each sequence is present in several copies per genome and absent in the genomes of two other amphibian species (Flavell et al., 1995).

REPTILIAN RETROTRANSPOSONS:

The genomes of snakes and iguanas contain retrotransposon. A 5' truncated Bov-B long interspersed repeated DNA (LINE) sequence has been identified in the genome of Vipera ammodytes. The phylogenetic distribution of Bov-B LINE among vertebrate classes show that besides the ruminantia, it is limited to Viperida snakes (Vipera ammodytes, Vipera palastinae, Echis coloratus, Bothrops alternatus, Trimeresurus flaviridae and T.gramineus. The copy number of 3'end of Bov-B LINE in the Vipera ammodytes genome is between 62000 and 75000 (Kordis and Gubensek, 1997). From three other reptiles, Conolopus subscristatus, Amblyrynchus cristatus and Pytas mucosus, Ty1 group retrotransposons have been isolated. The C. subcristatus sequence Tcs1 is present in multiple copies in both its host genome (Galapagos land iguana) and genome of the related Galapagos marine iguana (A. cristatus). Tpm1 has been isolated from the snake species, Pytas mucosus. (Flavell et al., 1995).

PISCLEAN RETROTRANSPSONS:

The element Easel has been isolated and characterized from several members of the piscine family salmonidae. It is one of the first LTR retrotransposons to be discovered in vertebrates. Phylogenetic analysis of this retroelement indicates that is belongs to gypsy group of LTR retrotransposons. (Tristem et al., 1995). From zebrafish (Danio rerio), a SINE called DANA has been isolated which exhibits all
the hallmarks of a SINE. DANA related sequences comprise approximately 10% of the modern zebrafish genome and some of the sequences containing full-length elements are capable of movement (Izsvak et al., 1996).

RETROTRANSPOSONS FROM SLIME MOLDS (Myxomycetes):
Among slime molds, retrotransposons have been identified from *Physarum polycephalum* and *Dictyostelium discoideum*. The repetitive fraction of the genome of slime mould *P. polycephalum* is dominated by *Tp1*, a family of retrotransposon-like sequences. *Tp1* elements consist of two terminal direct repeats of 277 bp which flank an internal domain of 8.3 kb (Rothnie et al., 1991). *Tp1* elements are arranged in scrambled clusters probably arising from integration of the element into copies of its own sequences. They are the major sequence component of hypermethylated (M+) fraction. *Tp1* elements are more closely related to *copia* and *Ty* elements. A second family called *Tp2* has also been described from the genome of *P. polycephalum*. Like *Tp1*, the *Tp2* element is structurally related to retrotransposons having LTRs and being flanked by an apparent target site duplication (Mucurrah et al., 1990) DIRS-1 element of *Dictyostelium discoideum* are unique among retrotransposons in being flanked by non identical long terminal inverted repeats (Boeke and Corces, 1989).

TRYPANOSOMA RETROTRANSPOSONS:
Several transposon like elements have been reported from a variety of trypanosome species. The genome of *Crithidia fasciculata* harbours two site-specific LINE like elements *CRE1* and *CRE2*. Studies of virus like particles (VLPs) produced in yeast cells after the over expression of fusion product of *Ty1* and *CRI* ORF have demonstrated that *CRE1* encodes an RNA-directed DNA polymerase. *CRE2* is 9595 bp element and is flanked by 29 bp target site duplications but unlike most other non-LTR retrotransposons lack 3' poly dA tract (Gabriel and Boeke 1991; Teng et al., 1995). The TRS/ingi elements are retrotransposon-like and present in many
copies in the genome of trypanosomes (Braun et al., 1992). These elements are highly repeated and have very long ORFs. One particular derivative of TRS/ingi, called TUBIS (truncated and inactivated form of ingi) had been found to interrupt a tubulin gene in a strain of *Trypanosoma brucei*. Both elements were shown to be present in several strains and species of the subgenus Trypanzoon, in particular in *T. brucei brucei, T. gambiense, T. rhodesiense, T. equiperdum, T. evansi*. Another elements named CZAR (cruzi associated retrotransposon) (Villanueva et al., 1991) and LITc are present in the genome of *T. cruzi* (Olivares et al., 1997).

**ANNELID RETROTRANSPOSONS:**

Annelids contain both LTR and non-LTR retrotransposons. Tas OE3, was cloned from the genome of parasitic nematode, *Ascaris lumbricoides*. The element is flanked by LTRs and contains three distinct regions encoding putative proteins typical for retroid elements (Felder et al., 1994). A sequence specific insertion named R4 present in the 26S rRNA gene of *Ascaris*, is shown to be a non-LTR retrotransposable element (Burke et al., 1995). The genome of *Caenorhabditis elegans* contains Ty3/gypsy class retrotransposon called Cer1. It is 8865 bp in length and has LTRs of 492 bp each. There is an exceptionally long (6819 bp) ORF (Britten, 1995). A family of non-LTR retrotransposons, called SR1 family, has been reported to be present in the genome of human blood fluke, *Schistosoma mansoni*. There are at least 200 copies of SR1 interspersed through the genome of *S. mansoni*. The structural and amino acid sequence similarities of SR1 with the members of CR1-like elements suggests that the SR1 family belongs to the CR1-like category of non-LTR retrotransposons (Drew and Brindley, 1997).