CHAPTER 2:
REVIEW OF LITERATURE
Transposable elements (TEs) are nucleic acid sequences that can insert into new chromosomal locations and often make duplicate copies of themselves in the process. With the advent of large-scale DNA sequencing, it has become apparent that, far from being a rare component of some genomes, TEs are the largest component of the genetic material of most eukaryotes. The active elements comprise only a tiny fraction of the TE complement in the genomes of maize and of most other multicellular organisms. However, the genomes of higher eukaryotes are filled with thousands, even millions, of TEs. Eukaryotic TEs are divided into two classes according to whether their transposition intermediate is RNA (class I) or DNA (class II). For all class I elements, it is the element’s transcript (mRNA), and not the element itself (as with class II elements), that forms the transposition intermediate. Each group of TEs contains autonomous and non-autonomous elements. Autonomous elements have open reading frames that encode the products required for transposition. Non-autonomous elements that are able to transpose have no significant coding capacity but retain the cis-sequences necessary for transposition. The transposition of some non-autonomous elements occurs with the help of autonomous elements.

2.1 Plant Retrotransposons

The class I elements or retrotransposons are ubiquitous throughout the plant kingdom and constitute a major portion of the nuclear genomes (in some cases as high as 50-70% of the genome) of plants (Kumar and Bennetzen, 1999). They are distributed as interspersed repeats almost throughout the length of all the chromosomes. Transposable elements were first discovered in plants but discovery of retroelements in plants started after Drosophila (e.g. copia and gypsy) and yeast (e.g. Tyl and Ty3), (Shepherd et al., 1984). Retrotransposons are indeed the most abundant and widespread class of transposable elements in plants (Table 1). On
the basis of presence or absence of LTRs, they are classified as LTR or non-LTR retroelements. Among the LTR retrotransposons, the copia and gypsy groups are widely distributed in the plant kingdom and are commonly found in high copy numbers (up to a million copies per haploid genome at least in plants with large genomes). The non-LTR retrotransposons, LINEs and SINEs, have also been found in high copy numbers, up to 250,000, in plant species studied so far. Recently sequenced Arabidopsis genome reveals that despite its small size, it has as many as 2109-class I elements comprising all the categories of retroelements, namely copia and gypsy types of LTR retrotransposons, non-LTR retrotransposons, LINEs and SINEs (Arabidopsis Genome Initiative, 2000).

Despite the slow research in this field, a large number of different plant retro-elements have been and are being discovered either by analyzing the insertions in or near various genes (Pelisier et al., 1995). Most of retro-elements identified are of Ty1-copia group, because of frequent use of Ty1-copia specific primers to amplify and clone Ty1-copia relatives (Flavell et al., 1992). However, Ty3-gypsy group of retrotransposons are also well represented. Some examples of the Ty1-copia group of retrotransposons are BARE-1 (barley), Bs1, Opie, PREM-1 (maize), SIRE-1 (soybean), Tnt1, Tto1 (tobacco), Tos17 (rice), OARE1 (Oat, Kimura et al., 2001), panzee (pigeonpea, Lall et al., 2002); and those of the Ty3-gypsy family are Athila (Arabidopsis), cereba (barely), cinful (barley), Grande-1, Zeon (maize), Magellan (maize, Michael et al., 1994), RIRE-3, RIRE8A, RIRE8B (rice, Kumekawa et al., 1999b), cyclops-2 (pea, Chavanne et al., 1998), Ogre (pea, Neumann et al., 2003), Gret1 (grape, Kobayashi et al., 2004), REM1 (Chlamydomonas, Perez-Alegre et al., 2005), LORE1 (Lotus japonicus, Madsen et al., 2005), Diaspora (soybean, Yano et al., 2005), PIGY (mung, Neumann et al., 2005) etc. In some gypsy-like retrotransposons (Athila, Cyclops-2, PIGY) an additional ORF down stream of the gag-pol region is present which has been hypothesized to be equivalent to the envelope gene of retroviruses that is why gypsy retrotransposons are said to be defunctionate retroviruses. The non-LTR
retro-elements LINEs and SINEs also have been identified in plants (Manninen and Schulman, 1993). Recently, terminal-repeat retro-transposons in miniature (TRIM) have been uncovered in Arabidopsis (Arabidopsis Genome Initiative, 2000) and apple (Antonius-Klemola et al., 2006). Initially these were observed during sequence analysis of a genomic clone containing the potato urease gene. TRIMs are said to be involved in restructuring plant genomes (Kumar and Hirochika, 2001).

2.2 Families and Structure of Retroelements

Members of retroelement family that contain reverse transcriptase encoding gene, responsible for the reverse transcription of DNA molecule from the RNA template, show high degree of variations. This reverse transcriptase or RNA dependent DNA polymerase was discovered over thirty three years ago as a retroviral encoded enzyme catalyzing DNA synthesis from RNA template (Baltimore, 1970; Temin and Mizutami, 1970). Since 1970s, different types of genetic elements from various organisms have been discovered. These genetic elements are shown to have open reading frames (ORFs) encoding sequences similar to retroviral reverse transcriptases (Temin and Mizutami, 1970). These retroelements fall into two main groups in plants: (i) transposable elements which like retro-viruses contain gag and pol genes along with long terminal repeats (LTRs); and (ii) without LTRs but have retroviral like gag and pol genes (Fig. 1). Similarly, in amino acid sequences the reverse transcriptases of these elements suggests a common origin for many diverse reverse transcriptase sequences (Xiong and Ecibush, 1990). Also, there exists sequence similarities among other coding regions but 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. Thus relationship between various retroelements has been established using the RT domain.
Figure 1. General structure of Ty1-copia, Ty3-gypsy, retrovirus, LINE and SINE retrotransposons
The genes within the retrotransposons encode capsid-like proteins (CP), endonuclease (EN), integrase (INT), protease (PR), reverse transcriptase (RT), envelop, and RNase H. PBS (primer binding sites), PPT (polypurine tracts). NA (nucleic acid binding moiety), IR (inverted terminal repeats), DR (flanking Target direct repeat), 5'UTR (5' untranslated region), 3' (3' untranslated region), and Pol IIIA and B-Promoter recognition sites for RNA polymerase III.
LTR retrotransposons contain direct long terminal repeats of generally 200-500 nucleotides that flank a region of 4 kb to 9.0 kb. The sequence divergence between the two LTRs of an element indicates the time of insertion of a particular retrotransposon, which were identical in the beginning. Most of the sequences display a high degree of variability (Casacuberta et al., 1995; Wessler, 1998) and in most cases, they represent elements that had lost the ability to transpose. This has been probably a consequence of intra-element LTR recombination, where the internal region has been lost (Vicient et al., 1999). The middle sequences contain one or more open reading frames (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 (Grandbastien et al., 1989). Both the gag and pol ORFs encode a polyprotein that is later cleaved into low-molecular weight individual functional proteins. The polyprotein encoded by pol is processed into (i) reverse transcriptase, which reverse transcribes RNA into double stranded DNA; (ii) aspartic protease, which cleaves the polyprotein into their component proteins (sometimes encoded by the gag gene); and (iii) RNaseH, responsible for RNA template degradation for insertion of double stranded DNA copy of the element into the host genome.

Retroelements with putative envelope genes have recently been described in several plants. The 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, but sometimes env-like ORF being present may encode analogous membrane spanning proteins, leading to doubt about infectious nature of retrotransposons (Granalbastein, 1992). This env ORF is well represented in Sireviruses, a plant-specific lineage of Tyl/copia elements (maize, Havecker et al., 2005), Diaspora of soybean (Yano et al., 2005), PIGY, a member of metaviridae family (Neumann et al., 2005), Athila (Voytas and Naylor, 1998) and Cyclops-2 of pea (Chavanne et al., 1998).
The structure of retrotransposons varies much in plants. In some plant retrotransposon families solo LTRs derived from unequal recombination between LTRs of a single element have been observed. Multiple nucleotide substitutions and small (one to four base pairs) insertions or deletions are frequently observed. Large internal rearrangements are also reported (Bennetzen, 1996). The presence of unrelated internal sequences in many related retrotransposons suggests that some of these are acquired sequences from other sources (Palmgren, 1994). Also, there exists a significant heterogeneity between LTRs of two related elements, which is thought to be responsible for faster evolution in retrosequences than host plant genome sequences.

The LTR retrotransposons and retroviruses have some similar steps in their life cycles. An integrated genomic copy of the element is transcribed by the cellular RNA polymerases, synthesizing mRNA encoding the proteins necessary for transposition and the template to synthesize a cDNA copy of the element. The mRNA is primed for reverse transcription by a specific cellular tRNA, which is reverse transcribed into double stranded cDNA by reverse transcriptase. Insertion of DNA copy occurs in presence of integrase. The integration of the element is always flanked by small target site duplication. Unlike DNA transposable elements that use a ‘cut-and-paste’ mechanism to move, RNA-mediated elements are necessarily replicate because the original copy, from which the mRNA has been transcribed, is not mobilized in the transposition process i.e., they follow ‘copy-and-paste’ mechanism. Integrated copies can be lost, however, through an unrelated mechanism involving recombination between the LTRs (see review, Feschotte et al., 2002).

The non-LTR retrotransposons are also found to make another large group, sometimes referred to as retroposon group. These non-LTR elements contain reverse transcriptase domain but are lacking aspartic protease and integrase region. In some cases gag like gene is also missing. Retroposons have a range of insertion site preferences; apparently some elements can insert randomly throughout the
genome, whereas others are restricted in their insertion to a single nucleotide sequence of a particular gene. Their mechanism of reverse transcription and integration is quite distinct from that of the LTR retrotransposons and is facilitated by an endonuclease (some times referred as integrase, it is not homologous to the LTR retrotransposon and retroviral integrases). The details of this mechanism have been described for R2 elements in the silkworm moth, *Bombyx mori*, which involves nicking of the target DNA and use of the exposed 3' hydroxyl group 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 an endonuclease with far less target site specificity than that of retroelement enzymes (Kumar and Bennetzen, 1999).

The SINE's comprise a class of retroelements distinct from LTR retrotransposons and non-LTR retrotransposons. Not only they are short (ranging from 75 to 500 bp in length), but also generally lack open reading frames, so can not code for specific enzymes (transposases) responsible for insertion process. So, they rather employ cellular mechanisms for retrotransposition. Research work on vertebrates (SINEs account upto 5% of the genome) has revealed 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' ends are quite variable in length, exact
sequences and also constitute the region of more heterogeneity among members of any given family. The A-rich 3’ end in quite variable in length and exact sequence and also 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 ‘A’s. In fact, simple sequences 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 repeat that flank the SINEs are not a part of repeated DNA family member itself, but derived 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 members of some families are not flanked by obvious direct repeats (Review, Bennetzen, 2000; Feschotte et al., 2002).

Apart from LTR and non-LTR retrotransposons, tyrosine recombinase encoding retroelements constitute a third group of retrotransposons. This group is represented by VIPER, Ngaro and DIRS elements (Lorenzi et al., 2005).

2.3 Mechanism of Retrotransposition and Retrotransposon Integration

It is observed that LTR retrotransposons use the same basic mechanism of retrotransposition in every system in which they exist including plants (Fig. 2A and B). Their basic mechanism of retrotransposition involves transcription of integrated element into a full length RNA, which is inserted into self encoded virus like particle (Bennetzen, 2000). The RNA is reverse transcribed into extrachromosomal DNA prior to insertion into genome by reverse transcriptase (Fig. 2A). Transcription of LTR retrotransposons does not necessarily correlate with new insertions in the genome. Thus, replication cycle for LTR retrotransposons includes four steps: transcription, translation, reverse-transcription and integration of element cDNA (Grandbastien, 1992; Bennetzen,
Figure 2(A). Mechanism of replication of retrotransposons. See text for description.
The replication is initiated by binding of tRNA to PBS and further extended towards 5’ end of 5’LTR by reverse transcriptase, when reached at the end of 5’LTR it is detached and get bound to R (repeat) region of 3’LTR and extended up to entire length towards 5’end. In this way both LTRs are maintained (Grice et al., 2003; Atwood-Moore et al., 2005). The sequences coding for gag and protease proteins are separated by several stop codons. The regions containing stop codons is removed from the element transcript by splicing and frame shifting (Neumann et al., 2003; Gao et al., 2003). Integration of retrotransposons is catalysed by integrase protein which is conserved among retroelements. Integrase binds retroelements and host DNA in a complex called preintegration complex (PIC) with other cellular proteins. Although no specific integration site is reported but the PIC is capable of directing the integration of retroelement DNA into any chromosomal location (Fig. 3A; Lewinski and Bushman, 2005). Such insertion specificity is associated with TIR transposable elements in all species investigated, including, plants where they are found to integrate preferentially into or near genes (Cresse et al., 1995; Peterson-Burch et al., 2004). The degree of target site selectivity varies among different retroelements. Some can apparently integrate at any location in the host genome, while others integrate almost at unique set of sequences. In case of animals, fungi, and retroviruses, retrotransposons show a strong preference for particular insertion sites. The yeast Tyl and Ty3 elements insert preferentially into regions upstream of tRNA genes and further enhanced by Set3 complex (histone deacetylase, Hos2 + trithorax-group protein, Set3), while TART of Drosophila specifically integrates at chromosome termini (Mou et al., 2006; Beliakova-Bethall et al., 2006)). Gel blot and in situ hybridization analyses have indicated that many retrotransposons are dispersed throughout their host genome except in centromeric regions (Bennetzen, 1996). Tos17 was shown to prefer gene-dense regions over centromeric heterochromatin regions, about 76% insertion are in genic regions (Miyo et al., 2003). Retroelements located in an intron are more often oriented in the direction opposite to that of the host gene.
Figure 2(B). Mechanism of Replication of retrotransposons. Left part. In retroviruses and retroelements that will form long terminal repeats (LTRs), reverse transcription begins near the 5' end using a tRNA hybridized to a region of the viral genome called the primer binding site (PBS) and soon reaches this end of the RNA molecule. The presence of the repeated sequence R (in yellow) at both ends of the RNA allows transfer of the nascent DNA to the 3' end of the genome (strong-stop strand transfer). The integrated form of the virus (provirus) is flanked by the LTRs. These are generated during reverse transcription, and contain the repeated sequence R and the unique sequences adjacent to it on the viral RNA (U5 in grey, and U3 in green, respectively). Synthesis of the new genomic RNAs will occur from a promoter in the U3 region. (Dark blue, cellular genomic DNA (the poly-A sequence is drawn as three As at the 3' end of the RNA)). Right part. By contrast, non-LTR elements begin reverse transcription near the 3' end of their RNA (using various primers) and do not require strand transfer. The example in the drawing represents replication of the human long interspersed element (LINE). Synthesis of the RNA will occur from an internal promoter located at the 5' end of the element.
Also transcription termination poly(A) signal motifs within retroelements interfere with normal gene transcription and are responsible for premature termination of host gene transcription (Cutter et al., 2005).

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 Cinful1) 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 retro-elements are abundant, in which they appear to be highly methylated and presumably heterochromatic, however, density centrifugation studies suggest that Cin4, Ta1 and Tnt1 are like Ac and Mu1 in their preferential association with genes containing isochors (Bennetzen, 1996). Recently inserted LTR retrotansposons are significantly more likely to be located in or within genes than are older, fragmented LTR retrotransposon sequences, indicating that most LTR retrotransponson-gene sequence associations are selected against over evolutionary time. LTR retrotransposon sequences (LRSs) associated with conserved genes are especially prone to negative selection. In contrast, fragmented LRSs that have persisted in the genome over long spans of evolutionary time preferentially associated with genes involved in signal transduction and other newly evolved functions (Ganko et al., 2006).

Plant retroelements are also found in mitochondrial DNA of Arabidopsis especially Ty1-copia and Ty3 gypsy group which were earlier thought to be associated with nuclear genome only (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. Sequence
Figure 3A. Variation of the copy number of the TAA repeats

(A) The I factor encoded endonuclease cleaves chromosomal DNA at the target for integration (arrows). (B) The RNA transposition intermediate which terminates downstream of the UAA repeats of the I element associates with the target DNA. (C) Reverse transcription initiates using the chromosomal DNA as primer and RNA as template, starting within the UAA repeats. (D) The RNA template dissociates from the chromosomal DNA and anneals to the newly synthesized sequence. (E) Reverse transcription extends into the I factor to complete first DNA strand synthesis. (F) After degradation of the RNA by RNAse H, second strand synthesis is completed. This process produces a new I element with an additional TAA repeat at the 3' end. The target DNA is in black, the RNA transposition intermediate in red with the extra-nucleotides flanking the I element in green, and the cDNA in blue. The target site duplication is underlined.
analysis shows that 5% of the \textit{Arabidopsis} mitochondrial genome appears to have retro-elemental origin.

Like those of other eukaryotes, fungal retrotransposons also transpose via RNA intermediate and employ reverse transcriptase for this purpose (Kempken and Kuck, 1998). In case of LINE1 dissemination is initiated by transcription from species-specific promoter located within 5'-untranslated region of LINE1 which is induced by binding sp1 and sp2 proteins specifically to the rat LINE1 promoter (Fedorov et al., 2005). Moran, (2002) reported that LINE-1 (L1) retrotransposition causes genomic deletions. Many times it results in the formation of chimeric L1s, containing the 5' end of an endogenous L1. Thus a number of studies demonstrate multiple pathways for L1 integration or retrotransposition (Moran et al., 2002). Regulation of replication cycle of LTR retrotransposons at any step can limit the transposition rate. For example, transcripts of the yeast \textit{Ty1} retrotransposon are abundant, but new insertions are extremely rare, largely because only one \textit{Ty1} cDNA, on average, is made for every 14,000 \textit{Ty1} transcripts (Curcio and Garfinkel, 1999).

\textbf{2.4 C-value Paradox, Diversity, Ubiquity and Copy number of retroelements}

The diversity of retrotransposons can be explained only if they belong to different individuals because of their large copy number. The retrotransposons isolated from individual plants are usually diverse at DNA sequence level, far more than similar retrotransposons amplified from \textit{Drosophila} and yeast (Flavell et al., 1992). Even the extremely small \textit{Ta} family of \textit{copia}-like retrotransposons of \textit{Arabidopsis} is highly diverse (Flavell, 1992).

Retrotransposons insertion in the 5' and 3' regions, near normal genes, coupled with their degenerate nature provides additional evidence for an ancient association between \textit{copia}-like retrotransposons and plant genomes. Maize gene \textit{adhl} present in 280 kb comprise of at least 37 classes of repetitive 197 kb DNA
sequences, a significant fraction of which are thought to be of retrotransposon origin (Avramova et al., 1995).

They are generally dispersed over plant chromosomes, consistent with their mode of amplification, but may associate with particular genomic regions. Most frequently, the rDNA and centromeric regions, consisting of tandemly repeated DNA elements, show a lower proportion of gypsy and copia like retroelements than do other regions (Kamm et al., 1996; Heslop-Harrison et al., 1997; Kubis et al., 1998a; Schmidt, 1999; McCarthy et al., 2002; Kobayashi et al., 2004). It is hypothesized that retroelements are more abundant around the centromeres of Arabidopsis chromosomes so as to limit the disruption of genes (Brandes et al., 1997). Relatively little is known about the chromosomal organization of LINEs. Retroelements due to their characteristic of automatic insertion into the genome provides putative source of biodiversity (Hirochika et al., 1996; Heslop-Harrison et al., 1997; Ellis et al, 1998; Flavell et al., 1998) and also markers of diversity.

A very large range of copy numbers from single copies of Ta1 in Arabidopsis and of Tst1 in Solanum tuberosum to >40,000 of del1 in Lilum longiflorum is observed (Grandbastien, 1992). Many other retrotransposons have been identified as extremely abundant sequences in plants (Feschotte et al., 2002; Tahara et al., 2004). These studies indicate that the large size of 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 del1 are present per haploid genome and they account for significant fraction of their host nuclear genomes (Bennetzen, 1996). The 4.45 kb LINE del2 has 250,000 copies in Lilum speciosum comprising 4% of its genome (Wessler et al., 1995). Extremely heterogenous populations of various retroelements have been shown to exist in Vicia sp. (Hill et al., 2005), mungbean (Galindo et al., 2004; Dixit et al., 2006), Sorghum (Muthukumar and Bennetzen, 2004), cotton (Zaki and Ghany, 2003; Zaki, 2005), Brassica (Alix and Heslop-Harrison, 2004), sweet potato (Tahara et al., 2004), sunflower (Natali et al., 2006) and within several other genera of higher
plants studied (Konieczny et al., 1991; Flavell et al., 1992; Hirochika and Hirochika, 1993; Vanderwiel, 1993; Pearce et al., 1996).

An important series of recent studies has shown that differential amplification of LTR-retrotransposons largely accounts for the 'C-value paradox' among the agronomically important members of the grasses. The C-value (the total amount of DNA in haploid genome) paradox may be defined as the lack of correlation between genomic complexity (C-value) to functional complexity of organisms (e.g., amphibians vs. mammals) i.e. genome size of an organism does not depend on its phylum (Feschotte et al., 2002). This paradox has been found applicable for both plant and animal species, but so far seems to be solved only for members of the grass family. In grasses the fraction of retrotransposons in the genome increases with its sizes, as is evident from rice, the smallest grass genome characterized (~389 Mb) consist of ~17% LTR retrotransposons (Jiang and wessler, 2001; McCarthy et al., 2002), maize with 50-80% retrotransposons of ~2,800 Mb genome (Bennetzen et al., 1998) (Morgante et al., 2001), barley with > 70% retrotransposons of ~4800 Mb genome (Vicient et al., 1999), cabbage with ~28% retrotransposons of ~600 Mb genome (Zhang et al., 2004) and *Vicia* species with ~45% retrotransposons of ~1300 Mb genome (Hill et al., 2005). Even species from the same plant family can exhibit striking differences in genome size, although the total number of genes might not be substantially different between them. The LTR-retrotransposons are mainly responsible for the vast differences in genome sizes between plant species (Bennetzen, 2000). In spite of a small genome size, the *Arabidopsis* genome harbour nearly 7% retroelement sequences (*Arabidopsis* Genome Initiative, 2000). Similarly, study of a 211 kb contiguous genomic region of *Triticum monococcum* orthologus to the Lr10 leaf rust resistance locus in hexaploid wheat reveals an overall gene density of one gene per 42 kb and 70% of these sequences comprise several classes of transposons and retro-elements (Wicker et al., 2001).
2.5 Why Retrotransposons are inactive during Normal Plant Development?

Many LTR-retrotransposons in plants appear to be defective as well, existing as solo LTRs or with internal deletions, rearrangements and/or replacements (Messing et al., 1995). This predominance of defective elements is partly due to the self-mutagenic properties of the retroelements to regulate their activity (Kunze and Becker, 1996; Grandbastien et al., 2005), but is also likely to be associated with an intrinsic higher mutation rate of cytosine-methylated DNA (Liu and Windel, 2000). It is likely that TEs methylation is associated with an inactive state of the chromatin, although it is not clear whether the methylation or chromatin alteration occurs first.

Retrotransposon inactivation may be explained on the basis of insertion. After insertion, transposase genes usually evolve as pseudogenes, quickly accumulating substitutions and insertions or deletions. Substitutions might introduce amino acid changes or stop codons. This can result in a disrupted open reading frame (ORF) and/or an inactive transposase (Arabidopsis Genome Initiative, 2000). The precise mechanism(s) of epigenetic regulation of transposable elements remain(s) unclear but the phenomenon does have similarities with the homology-based silencing that has been observed with plant transgenes. In fact, it is highly likely that the transgene silencing process is a secondary outcome of an evolved plant mechanism for the inactivation of plant viruses and TEs. Wide crosses can reactivate silenced transposable elements in Drosophila and a large amplification of genome size associated with a wide cross has also been observed in Wallabies (Graves et al., 1998).

The study of transgene silencing and viral resistance in plants and of TEs regulation in animals led to the identification of two distinct epigenetic mechanisms known as post-transcriptional gene silencing (PTGS) and transcriptional gene silencing (TGS). In PTGS (a process also referred to as co-suppression and related to RNA interference in animals), TEs silencing is caused
by degradation of their RNAs. In TGS, TEs are transcriptionally repressed (Hirochika et al., 2000; Wright and Voytas, 2002; Aravin, 2001; Djikeng et al., 2001; Lindroth, 2001; Steimer, 2000). Although PTGS has been well documented in plants, most notably as a defense against viral replication, TGS seems to be the principal pathway to silence plant transposable elements.

In all cases, genetically inactive elements were hyper-methylated (especially at their termini, where the transposase promoter resides in the autonomous family members) relative to their active counterparts whereas hypomethylation was found to be a hallmark of actively transcribed and transposing elements (Liu et al., 2004). Hypermethylation is also associated with the intergenic clusters of LTR-retrotransposons in maize. These regions, which make up at least 50% of the maize genome, are highly condensed and thought to comprise a transcriptionally repressive chromatin environment (Meyers et al., 2001; SanMiguel et al., 1996).

As mentioned above the methylation of TEs sequences (especially the promoter of transposase genes) correlates with TEs inactivation in maize. Similarly, the methylation of transgene promoter sequences correlates with TGS in plants (Vaucheret et al., 2001). Therefore, it is not surprising that endogenous TEs can be activated in mutants that are impaired for the establishment and maintenance of TGS. For example, in the Arabidopsis ddm1 (decrease DNA methylation) mutant, endogenous transposons of the Mutator and CACTA (En/Spm) super families of TEs are transcriptionally and transpositionally reactivated (Singer et al., 2001; Miura, 2001). Plants that are homozygous for the ddm1 mutation have notably decreased CpG methylation. Consequently, the transcriptional derepression of TEs in the mutant strains was accompanied by demethylation of the elements (Kato et al., 2003). The ddm1 gene encodes a protein with strong similarity to SWI2/SNF2 chromatin-remodeling factors (Richards et al., 1999). These results indicate a possible functional link between chromatin remodeling, DNA methylation and genome integrity.
Reactivation of TEs *ddml* mutant background might have a direct impact on host fitness, as the reactivated TEs have been found to insert into coding and regulatory sequences of genes (Singer et al., 2001; Miura, 2001). It is likely that many of the developmental abnormalities that were observed in *ddml* plants were induced by the movement of reactivated transposable elements.

Recently it has been discovered that the mRNA initiation and features of primary and secondary structure of tobacco retrotransposon *Tto1* RNA influence its transpositional activity in heterologous host *Arabidopsis*. Surprisingly reverse transcription of most transcribed *Tto1* RNA does not occur but it is poorly translated, and its 5'-'end does not contain a region of redundancy with more prominent 3'-end. In contrast, expression of mRNA with the 5'-end extended by 28 nucleotides allows translation and gives rise to transposition events in the heterologous host. In addition, the presence of extended hairpins and two short ORFs in the 5'-leader sequence of *Tto1* mRNA suggest that translation does not involve ribosome scanning from the mRNA 5'-end to the translation initiation site (Bohmdorfer et al., 2005).

### 2.6 Expression and Activation of Retrotransposons

A large number of retrotransposons have been isolated from various plants, but there is little knowledge about their expression. 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 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 (Feschotte et al., 2002, review). In contrast, retrotranscripts can be quite abundant in yeast (Curcio and Garfinkel, 1999) and *Drosophila*. Initial studies of cloned retroelements from plants did not show 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; Bohmdorfer et al., 2005) 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 (Vicient et al., 2005). 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 (Bohmdorfer et al., 2005).

During study of expression of components of the yeast retrotransposon Ty1 in *E. coli*, it was found that polypeptides encompassing the capsid-forming component of Ty1 retrotransposon can assemble into particles in the heterologous host. Ty RNA can be detected in particle fractions. RNA packaging depends on features in the 5' part of Ty RNA, because deletion of 5' proximal sequences leads to decreased packaging efficiency. The observations support that these steps are independent of host factors (Luschnig and Bachmair, 1997). The gag homologue of Ty1 assembles into spherical particles similar, but not identical to virus like particles in the natural host of Ty1, *Saccharomyces cerevisiae*. Assembly process fails upon deletion of a domain in the C-terminus of reading frame of Ty1. Also, Ty1 gag fusion proteins can be produced in *E. coli* an organism lacking endogenous retrotransposon (Bachmair et al., 1995). It has been observed that telomere erosion triggers an increase in Ty1 cDNA in yeast (Scholes et al., 2003). Many retrotransposons show unique patterns of developmental and/or environmental regulation. This can be explained by the presence of limited data on the activity of plant LTR retrotransposons, indicating that transposition is regulated primarily at the level of transcription initiation (Grandbastein et al., 2001; Bohmdorfer et al., 2005). Many of the plant retrotransposons studied so far are transcriptionally activated by various biotic and abiotic stresses (Wessler, 1996; Mhiri et al., 1997; Beguiristain et al., 2001). Expression of the tobacco *Tnt1* and *Tto1* retrotransposons is greatly enhanced by several abiotic stresses including
Table 1. Active or potentially active plant retrotransposons

<table>
<thead>
<tr>
<th>Classification</th>
<th>Species</th>
<th>Evidence for recent mobility</th>
<th>Evidence for transcripts</th>
<th>Refs</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Subclass I: LTR retrotransposons</strong>&lt;br/&gt;Superfamily Ty1/copia</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BARE-I</td>
<td>Barley</td>
<td>None</td>
<td>In leaves and callus</td>
<td>Suoniemi A et al., 1996</td>
</tr>
<tr>
<td>B5</td>
<td>Maize</td>
<td>Transposition into the waxy gene</td>
<td>None</td>
<td>Varagona MJ et al., 1992</td>
</tr>
<tr>
<td>OARE1</td>
<td>Oat</td>
<td>Transcript detected</td>
<td>In leaves</td>
<td>Kimura Y et al., 2001</td>
</tr>
<tr>
<td>Hopscotch</td>
<td>Maize</td>
<td>Transposition into the waxy gene</td>
<td>None</td>
<td>White SE et al., 1994</td>
</tr>
<tr>
<td>Ji</td>
<td>Maize</td>
<td>None</td>
<td>In roots(^b), leaves(^b) and tassels(^b)</td>
<td>Avramova Z et al., 1995 &amp; SanMiguel P et al., 1996</td>
</tr>
<tr>
<td>Opie</td>
<td>Maize</td>
<td>None</td>
<td>In roots(^b), leaves(^b) and tassels(^b)</td>
<td>Avramova Z et al., 1995 &amp; SanMiguel P et al., 1996</td>
</tr>
<tr>
<td>Oser</td>
<td>Volvox carteri</td>
<td>Copy with identical LTRs</td>
<td>None</td>
<td>Lindauer A et al., 1993</td>
</tr>
<tr>
<td>PREM-2</td>
<td>Maize</td>
<td>None</td>
<td>In early microspores(^c)</td>
<td>Turcich MP et al., 1996</td>
</tr>
<tr>
<td>Prf1 and Prf2</td>
<td>Potato</td>
<td>None</td>
<td>In protoplasts(^b)</td>
<td>Pearce SR et al., 1996</td>
</tr>
<tr>
<td>Prf3</td>
<td>Potato</td>
<td>None</td>
<td>In protoplasts(^b)</td>
<td>Pearce SR et al., 1996</td>
</tr>
<tr>
<td>R9'</td>
<td>Rye</td>
<td>None</td>
<td>In seedlings(^b)</td>
<td>Pearce SR et al., 1997</td>
</tr>
<tr>
<td>SIRE-1</td>
<td>Soybean</td>
<td>None</td>
<td>In seedlings and leaf tissues</td>
<td>Bi Y-A et al., 1996 &amp; Laten H, Unpublished</td>
</tr>
<tr>
<td>Stonor</td>
<td>Maize</td>
<td>Transposition into the waxy gene</td>
<td>None</td>
<td>Varagona MJ et al., 1992 &amp; Marillonnet S, Unpublished</td>
</tr>
<tr>
<td>Tnp2/Tnt1B</td>
<td>Nicotiana plumbaginifolia</td>
<td>Transposition into the nia gene in protoplast cultures</td>
<td>None</td>
<td>Meyer C, Unpublished</td>
</tr>
<tr>
<td>Tnt1A</td>
<td>Tobacco</td>
<td>Transposition into the nia gene in protoplast cultures; small increase in copy number in cell cultures</td>
<td>In roots, in protoplasts, and after wounding and pathogen attacks</td>
<td>Grandbastien M-A et al., 1997</td>
</tr>
<tr>
<td>Tto17</td>
<td>Rice</td>
<td>Copy number increase and active transposition into coding sequences in cell and tissue cultures</td>
<td>In cell cultures</td>
<td>Hirochika H et al., 1996</td>
</tr>
<tr>
<td>Tto19(^f)</td>
<td>Rice</td>
<td>Small increase in copy number in cell cultures</td>
<td>In cell cultures(^b)</td>
<td>Hirochika H et al., 1996</td>
</tr>
<tr>
<td>Tto1</td>
<td>Tobacco</td>
<td>Copy number increase in cell and tissue cultures; transposition into the nia gene in protoplast cultures</td>
<td>In protoplasts, cell and tissue cultures, and after wounding and viral attack</td>
<td>Hirochika H, 1993, Takeda S et al., 1998 &amp; Grappin P et al. Unpublished</td>
</tr>
<tr>
<td>Tto2c</td>
<td>Tobacco</td>
<td>Small increase in copy number in cell cultures</td>
<td>In protoplasts(^b)</td>
<td>Hirochika H, 1993</td>
</tr>
<tr>
<td>Tto5c</td>
<td>Tobacco</td>
<td>None</td>
<td>After viral attack(^b)</td>
<td>Hirochika H, 1995</td>
</tr>
<tr>
<td>Wis-2</td>
<td>Wheat</td>
<td>Polymorphism in regenerated plants</td>
<td>In protoplasts</td>
<td>Moore G et al., 1991 &amp; Lucas H, Unpublished</td>
</tr>
<tr>
<td><strong>Superfamily Ty3/gypsy</strong>&lt;br/&gt;</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cinful</td>
<td>Barley</td>
<td>None</td>
<td>In leaves(^b)</td>
<td>Avramova Z et al., 1995 &amp; SanMiguel P et al., 1996</td>
</tr>
<tr>
<td>Huck</td>
<td>Maize</td>
<td>None</td>
<td>In roots, leaves and tassels</td>
<td>Avramova Z et al., 1995 &amp; SanMiguel P et al., 1996</td>
</tr>
<tr>
<td>magellan</td>
<td>Maize</td>
<td>Transposition into the waxy gene</td>
<td>None</td>
<td>Paragaman MD et al., 1994</td>
</tr>
<tr>
<td>TCI-4</td>
<td>Tomato</td>
<td>None</td>
<td>In seeds(^b)</td>
<td>Su P-Y et al., 1997</td>
</tr>
<tr>
<td>Zeon-1</td>
<td>Maize</td>
<td>Transposition near the zeinA gene in somatic tissues</td>
<td>In endosperm(^b)</td>
<td>Hu W et al., 1995</td>
</tr>
<tr>
<td>LORE1</td>
<td>L. japonicus</td>
<td>Transposition into genes</td>
<td>In roots</td>
<td>Madsen et al., 2005</td>
</tr>
<tr>
<td><strong>A typical or not yet classified</strong>&lt;br/&gt;</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Br1</td>
<td>Maize</td>
<td>Transposition into the Adh gene after viral infection</td>
<td>None</td>
<td>Johns MA et al., 1985</td>
</tr>
<tr>
<td>PREM-1</td>
<td>Maize</td>
<td>None</td>
<td>In early microspores(^b)</td>
<td>Turcich MP et al., 1994</td>
</tr>
<tr>
<td>TOC1</td>
<td>Chlamydomonas reinhardtii</td>
<td>Transposition into the OEE1 gene and increase in copy number during mitotic growth</td>
<td>Yes</td>
<td>Day A et al., 1991</td>
</tr>
<tr>
<td>Sln3</td>
<td>Rapseseed</td>
<td>None</td>
<td>In shoots, roots and callus</td>
<td>Deraon J-M et al., 1996</td>
</tr>
<tr>
<td>TS</td>
<td>Tobacco</td>
<td>None</td>
<td>By in vitro transcription</td>
<td>Yoshioka Y et al., 1993</td>
</tr>
</tbody>
</table>

\(^a\)The table includes all those plants retrotransposons for which transcripts or mobility have been reported. However, any transcripts only shown to be expressed from foreign promoters are excluded: elements showing intervarietal or interallelic polymorphisms, indicative of fairly recent activity, are also excluded, because this is poor criterion for present activity. \(^b\)The possibility that transcripts are initiated from external upstream promoters has not been ruled out. \(^c\)Partial sequences isolated by PCR methods. \(^d\)H. Laten, unpublished. \(^e\)S. Marillonnet and S. Wessler, unpublished. \(^f\)C. Meyer, unpublished. \(^g\)P. Grappin and M-A. Grandbastien, unpublished. \(^h\)H. Lucas, unpublished. \(^i\)C. Kinlaw, unpublished.
protoplast isolation, cell culture, wounding, methyl jasmonate, CuCl₂, UV and salicylic acid (Takeda et al., 1998, 1999; d’Erfurth et al., 2003a; Grandbastien et al., 2005; Iahizaki et al., 2005). The same is true with BARE-1 and OARE-1 retrotransposon of barley and oat respectively (Vicient et al., 1999; Schulman et al., 2000; Kimura et al., 2001). Similarly, biotic stresses, such as infection by viral, bacterial and fungal pathogens, can activate the transcription of these elements (Grandbastien et al., 2001), as shown of Tnt1 by fungal factors. 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). Transcription of Tnt1 elements is also induced when tobacco cells are treated with fungal extracts containing cell wall hydrolases (Pouteau et al., 1991). Certain biotic factors of microbial origin are also known to induce Tnt1 promoter only when the LTR promoter is placed in a heterologous context (Grandbastien et al., 2005).

In contrast to OARE1 and Tnt-1, transcription of Tos10, Tos17, Tos19 in rice and Tio1, Tio3, in tobacco is induced by cell culture and protoplast culture respectively (Hirochika, 1993; Hirochika et al., 1996). Treatment with 5-azacytidine induced both CG and CNG partial hypomethylation of Tos17 in rice (Liu et al., 2004). Plants regenerated from cultured tobacco or rice cells possess new retrotransposon insertions making it possible that activation of retrotransposon is responsible for somaclonal variations, that are stable (Wessler et al., 1995). After induced transcription, the genomic copy number of the rice LTR retrotransposon Tos17 increased from 2 to more than 30 copies in some strains (Hirochika et al., 1996; Grandbastin, 1998; Liu et al., 2004). Only a fraction of retrotransposon population in plants appears to be transcriptionally active as revealed by a recent survey of EST databases that indicates that approximately 1.2% of the total sequences represent retrotransposon complementary sequences (Vicinet et al., 2001). Partial transcripts of retrotransposons have been amplified from other plants like sweet potato (Tahara et al., 2004) and sorghum.
Review of Literature

(Muthukumar and Bennetzen, 2004). The transcript of rice gypsy-like retrotransposon RIRE9 includes the coding sequence for the C-terminal part of the RNaseH domain and the N-terminal part of the integrase domain in the polyprotein region (Liu et al., 2000). In case of pea retroelement Ogre the sequences coding for gag and protease proteins are separated from thr rt/rh-int domains by several stop codons. The region containing these stop codons is removed from the element transcript by splicing which results in reconstitution of the complete gag-pol sequence. Only a part of the transcripts is spliced which probably determines the ratio of translated proteins (Neumann et al., 2003)

There is strong evidence that retroelements may confer certain selective advantage on the system possibly to withstand adverse environmental stresses. Analysis of BARE1 element of barley in “Evolutionary Canyon” Mount Carmel in Israel indicates that plants grown at the top of the Canyon have three times more retroelements than the plants grown near the bottom of the Canyon. Plants grown at higher elevation apparently gained more copies of retroelements. Kalender et al., (2000) have speculated that a larger genome achieved through the ample presence of retrotransposons might help plants to cope up with more stressful high and dry areas of Canyon. This is consistent with the suggestion that sunlight, that is likely to be more plentiful at higher elevation of evolutionary Canyon, may be an important environmental agent involved in genomic restructuring (Moffat, 2000), as it has been shown earlier that shorter wavelength UV light can activate retrotransposons (Walbot, 2000; Kimura et al., 2001).

2.7 Origin of retro-elements

The presence of transposable elements in all living organisms and a very high level of sequence conservation of retroelements suggest an early origin of these mobile elements. The retroviruses can travel both within members of a species and interspecies efficiently. So their time of origin and specific mechanism
involved in origin can not be explained clearly. Since retrotransposons are the only elements common to both the LTR and non-LTR groups of the retroelement family, their structure is the most likely progenitor of all retroelements known today (Xiong and Eickbush, 1990). Vertebrate retroviruses appears to have more active RNaseH(RNH) domain, but other LTR-retrotransposons have acquired an enzymatically weak RNaseH(RNH) enzymes to ensure correct processing of the polypurine tract (PPT), which is an important step in the life cycle of these retrotransposons. The degeneration of RNH domains indicates that LTR-retrotransposons arose late in eukaryotes (Malik, 2005). It is also evident that most copies of retroelements do not occur in tandem arrays, although rare genomic clones sometimes carry two or more copies of the same element (Bennetzen, 1996). The barley BIS1 elements is lacking or reduced in centromeres telomeres and nucleolar organizer (Moore et al., 1991), while the Grandel elements of maize is observed to be under-represented at centromeres and somewhat clustered in the distal regions of some chromosome arms (Palmgren, 1994). Analysis of different SINEs indicates independent origins, from different RNA polymerase III products.

A large number of LTR-retrotransposons belonging to different families contribute over 70% of the maize nuclear genome. These retrotransposon sequences mostly appear to have arisen within the last 2 to 6 million years. It is possible that low copy numbers of these elements existed in the maize genome long before this time and their amplification was a recent event. Alternatively, they may have arisen via horizontal transfer within this short time duration, either as a naked nucleic acid or within a packed virus (Pardue et al., 1997). Each of major clades of rice LTR-retrotransposons is more closely related to elements present in other species than to the other clades of rice elements, suggesting that horizontal transfer may have occurred over the evolutionary history of rice retrotransposons (McCarthy et al., 2002). These types of transfers might have occurred into injured tissues. This model favours origin of retrotransposons in
Review of Literature

vegetatively reproducing plants only. The LTR-retrotransposons of maize share many properties with retroviruses, including their ability to acquire sequences from other genes (Kumar et al., 1996; Palmgren, 1994; Wessler et al., 1994). Like retroviruses, retrotransposons can acquire env encoding sequences (Messing et al., 1995; Gaucher et al., 1998; Yano et al., 2005). Thus, some defective retroviruses might have given rise to LTR-retrotransposons (Bennetzen, 2000) because defective appearance of the env-derived regions in LTR-retrotransposons has been reported in many organisms including plants such as Arabidopsis, pea and soybean (Gaucher et al., 1998; Laten et al., 2003; Lorenzi et al., 2006; Neumann et al., 2005; Yano et al., 2005).

2.8 Retrotransposons in Relation to Retroviruses

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. Retrotransposon mutation rate has been reported to be 1 in 7000 to 50,000 residues per replication. Thus even a few cycles of insertion, transcription, reverse transcription and reinsertion would lead to rapid divergence in the original sequence of a retrotransposon. In absence of any advantage the sequences could be lost. 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 (Jin and Bennetzen, 1989; Lewinski et al., 2005).

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 (Feschotte et al., 2002 review). So, the percent of divergence will indicate how long the element has been a resident within the genome.
Since retrotransposons use the same proteins for their replication as are used by some integrating retroviruses, it is thought that some retrotransposons are similar to integrated retroviruses (Grice et al., 2003). These forms of retrotransposons can cross the cellular barriers like retroviruses. Intracellular virus-like particles (VLP) have been observed for several LTR retrotransposons, but they lack (env) envelope protein coding gene necessary for intercellular transmission, But certain endoviruses also lack envelope coding gene. Diaspora, a soybean retroelement is the only fully characterized member of a lineage of putative plant endogenous retroviruses that contain no trace of an extra coding region. The loss of an envelope-like coding domain suggests that non-infectious retrotransposons could swiftly evolve from infectious retroviruses, possibly by anomalous splicing of genomic RNA (Yano et al., 2005). Envelope gene containing retroviruses-like elements such as Athila in Arabidopsis, PIGY in pea, SIRE1 in soybean are widely present and expressed (Laten et al., 2003; Neumann et al., 2005). Kim et al. (1994) reported gypsy infectious to Drosophila raised on a medium containing homogenized pupae from a gypsy active Drosophila line. Sequence comparisons of gypsy, copia and retroviruses showed a significant homology between gypsy and retroviruses than copia and retroviruses. Thus gypsy could be considered for infectious nature of some retrotransposons (Sinkovics, 2001). Vertebrate retroviruses appear to have active RNaseH domains, but LTR-retrotransposons have degenerated RNaseH domains (Malik, 2005). Both retrotransposons and retroviruses assemble intracellular immature core particles around a RNA genome, and nascent particles collect in association with membranes or as intracellular clusters. These virus-like particles and mRNA and proteins of yeast Ty3-gypsy retrotransposons accumulate in association with cytoplasmic P-bodies, which are sites of mRNA translation repression, storage and degradation (Beliakova-Bethell et al., 2006)

Retroviruses can acquire and transmit portion of their host genome by process of transduction. Viruses can take up protooncogenes into viral genomes and
convert them into oncogenes; also they can take up other cellular functions also. Acquired cellular genes have usually replaced essential viral genes thus inducing a functional virus into a defective one that requires a helper virus for infectivity. This proposed mechanism of host gene acquisition suggests that retroelements other than retroviruses should have ability to acquire the host genes, but this has not been observed for any animal or fungal retrotransposon. It is still disputable to say that the acquisition of nuclear genes is a unique feature of retroviruses. If it is, then elements such as Bs1 should be defective version 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. Sequence 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, 1989). Also Bs1 belongs to Ty3-gypsy groups of retrotransposon 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 conservative mutations since acquisition from Mha1, suggesting a selection for function. The portion of Mha1 acquired could specify attachment to plasma membrane, perhaps supplying an env function. Bennetzen et al (1996) has sequenced cinful-1 a gypsy group element which is first plant retroelement to contain coding potential for all intracellular functions and ordered additional sequences that could specify env proteins. Zeon-1 appears to be 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 sequences with no apparent sequence similarity to known retroelement gene. The possible origin of these internal sequences is still to be determined though it is hypothesized that they might originate from normal nuclear genes (Pelissier et al., 1995).
Although mutation has chaotic aspects, spontaneous mutation rates assume certain characteristic values when expressed per genome per genome duplication. The rate among lytic RNA viruses is roughly 1, while the rate among retroelements is roughly estimated to be 0.2. It is also suggested that retrotransposons and endogenous retroviruses might have emerged in theropod dinosaurs when Aves evolved, and directed the development of syncytiotrophoblasts in the placentae of the first mammals (Sinkovics, 2001). This is suggested on finding that RNA genomes derived from ancestors of viroids make ribozymal entry into vesicle containing autocatalytically replicating oligopeptides to bring about RNA proliferation and enzyme synthesis with in the vesicle (Sinkovisz, 2001).

2.9 Retroelements in Genome Evolution and Biodiversity

LTR-retrotransposon, which are located largely in intergenic regions are the single largest component of most plant genomes (Kumar and Bennetzen, 1999). LTR-retrotransposons were first discovered in plants as sources of both spontaneous and induced mutations in maize and tobacco (Johns et al., 1985; Grandbastein et al., 1989; Wessler et al., 1992). As with the active class II elements that are responsible for unstable mutations, the mutagenic LTR-retrotransposons are members of low to moderately repetitive element families (Hirochika et al., 1996; Wessler et al., 1998; Grandbastein et al., 1989; Meyers et al., 2000). For example, the Bs1 element, which was first detected as an insertion in the alcohol dehydrogenase1 gene (adhl), is present in only 1-5 copies in the maize genome (Johns et al., 1985). The interactions between retrotransposons and their host are the causes of the abundance of transposable elements in the genome that they occupy (Brookfield et al., 2005).

Due to their property of automatic insertion into the genome, retroelements act as mutagenic agents, thereby providing a putative source of biodiversity (Hirochika et al., 1996; Heslop-Harrison et al., 1997; Ellis et al., 1998; Flavell et
Review of Literature

al., 1998) and serving as molecular markers for estimation of biodiversity. Regulatory mechanism, may act to protect genomes from insertional mutagenesis (Lucas et al., 1995), and it has been suggested that transgene-induced silencing reflects mechanism aiming to prevent genome invasion by retroelements. Insertions of retrotransposons results into either inactivation or alteration in gene function (Wessler et al., 1995). Nearly 80% mutations detected in *Drosophila* are caused due to insertions. Transposons also can excise, partially or completely restoring gene function, and can also lead to chromosome rearrangements such inversions and translocations. In yeast various Ty retroelements are reported to act as agent of genome rearrangement primarily because they serve as sources of homology for ectopic (or unequal) recombination. Unequal recombinations between directly repeated elements at adjacent sites will give rise to reciprocal duplications and deletions of the DNA between the two elements, while unequal exchange between elements in opposite orientations will yield an inversion of DNA between the elements. Similar ectopic exchange between elements on different chromosomes can give rise to reciprocal translocation. All of these rearrangements and more complex events requiring more than one ectopic recombination events have been observed in yeast. In *Drosophila*, such unequal recombination events have also been observed both between the two LTRs of an LTR-retrotransposon to give a solo LTR and between two distant transposable elements. Such an equal recombination was the source of the first gene duplication event ever reported, generating the *Bar* eye phenotype in *Drosophila* (Sabl and Henikoff, 1996). In plants, as in other eukaryotes, most recombination is limited to genes, thus limited ectopic recombination has been reported (Bennetzen, 2000 review). In *Arabidopsis* constant turnover of euchromatic insertions and a decline in activity of the element that target heterochromatin have limited the contribution of retrotransposon DNA to the genome size expansion. Insertions tend to be lost from euchromatin regions where they are less likely to fix in comparison to heterochromatin regions (Pereira, 2004). In rice approximately 20% of the LTR-
retrotransposons sequences lie within putative genes. The distribution of elements across chromosome 10 is non-random with the highest density (48 element per Mb) being present in the pericentric region (Gao et al., 2004).

Transposable elements can also act to move elements such as exons and promoters into existing sequences so as to create new gene functions and contribute to evolution (Moran et al., 1999). Alternative splicing of genes caused by transposable elements has been shown in maize (Bureau and Wessler, 1994). The sequences of degenerate and potentially active retroelements can give valuable data about genome evolution and phylogenetic relationship if interpreted carefully. Although retroelement amplification leads to larger genomes (Bennetzen, 1996), it is probable that retroelement turnover and loss can occur in a directed manner (Tatout et al., 1998), leading to different retroelement compositions between species. For example, chromosome sets in the cultivated hexaploid oat, *Avena sativa*, can be discriminated by the presence of retroelement families (Katsiotis et al., 1996).

It has been known since the late 1980s, however, that both LTR and non-LTR retrotransposons can attain phenomenally high copy numbers in plant species that have large genomes (Kumar and Bennetzen, 1999). In three species in the *Vicia* genus *copia* retroelement copy number varies from 1000 to 1 million with more sequence heterogeneity being present in species with higher copy number (Pearce et al., 1996). Although in part due to random mutations of the high number of copies present in most plant genomes, sequence variability is often non-uniformly distributed along the retroelement: regulatory regions (including the LTRs of copia element) can evolve faster than coding region, perhaps enabling elements to coexist with their host genomes without detriment (Vernhettes et al., 1998). LTR-LTR recombination has generated tandem multimeric *BARE1* elements in which internal coding domains are interspersed with shared LTRs. The data indicates that recombination contributes to the complexity and plasticity of retroelement evolution in plant genomes (Vicient et al., 2005). The incredible
potential of amplification through retrotransposition in plants was first illustrated by studies on members of the genus *Lilium*. The 14 species that were surveyed have enormous genomes of 30,000-45,000 Mb, the size of which seems to result from massive amplification of retrotransposons (Yoshioka et al., 1993).

An important series of recent studies has shown that differential amplification of LTR-retrotransposons largely accounts for the 'C-value paradox' among the agronomically important members of the grasses. The C-value paradox is the observed lack of correlation between increases in DNA content and the complexity of an organism. This paradox has been documented for both plant and animal species, but so far seems to be solved only for members of the grass family. In this family, the fraction of the genome contributed by LTR retrotransposons increases with genome size from rice, the smallest characterized grass genome (~17% of its 389 Mb genome consist of LTR retrotransposons) (McCarthy et al., 2003), to maize (~2,800 Mb, 50-80% retrotransposons, SanMiguel et al., 1998; Meyers et al., 2001) and barley (~4,800 Mb, >70% retrotransposons, Vicient et al., 1999). In rice an average of 22.3 LTR-retrotransposons per Mb were detected in chromosome 10. *Gypsy*-type elements were found to be >4x more abundant than *copia* elements (Gao et al., 2004).

Some studies have raised the question on our concept of the dynamic genome concept and have positioned the grass clade as a focal point for future studies. In a classic study, Jeff Bennetzen et al., (1996) analyzed a 280 kb region around the maize adh1 gene and found that nested LTR retrotransposons accounted for most of this sequence. This clustering of LTR-retrotransposons in intergenic regions was shown to be representative of the rest of the genome. Their initial observations were dramatically followed up with the demonstration that bursts of LTR retrotransposon activity have doubled the maize genome within the past 6 million years (Bennetzen et al., 1996; SanMinguel et al., 1998). The temporal components to their analysis were made possible by exploiting the fact that the LTRs of a single element are identical on insertion. By comparing the two
LTR sequences of a single element, they were able to estimate the insertion time. At least 46.5% of LTR-retrotransposons in the rice genome are older than the age of the species (<680000 years; Gao et al., 2004). This result showed for the first time that TEs could rapidly restructure a genome. Recently a new group of long terminal repeats (LTR) retrotransposons, termed terminal repeat retrotransposons in miniature (TRIM) are reported to be involved in restructuring plant genomes (Witte et al., 2001).

In another study, Kalender et al., (2000) presented a striking example of transposable element-mediated genome restructuring in populations of the wild barley *Hordeum spontaneum*. In this case, genome restructuring takes the form of pronounced intra-specific genome size variation due to amplification of the *BARE*-1 LTR retrotransposon. The copy number of *BARE*-1 among nearby populations that are subjected to different levels of water stress varied between 8,300 and 22,100 corresponding to 1.8 to 4.7% of the nuclear DNA. The correlation between *BARE*-1 copy number, genome size and local environmental conditions indicated that a mechanistic connection might exist between the amplification of a particular transposable element and the adaptive evolution of its host (Grandbastien et al., 1998; Kalendar et al., 2000; Vicent et al., 2005).

Diverse and extreme heterogeneous population of LTR-retrotransposons has been predicted to exist in *Solanum sp.* (Flavell et al., 1992), oat (Kimura et al., 2001), cotton (Talierico et al., 2003; Zaki, 2005), common bean (Galindo et al., 2004), sorghum (Muthukumar and Bennetzen, 2004), *Brassica* (Alix et al., 2004, 2005), *Vicia sp.* (Hill et al., 2005), sunflower (Natali et al., 2006) and mungbean (Dixit et al., 2006).

### 2.10 Utility of retrotransposons as molecular markers

The success of molecular breeding hinges upon good genetic linkage mapping data and identification of markers closely linked to genes influencing important agronomic traits. DNA based markers have been proved important for
linkage and polymorphic studies. The DNA based markers are classified into two major categories based on their basic techniques: Hybridization based, and PCR based. The hybridization-based marker such as RFLP (Restriction Fragment Length Polymorphisms) requires prior sequence information for probe development and also screens very few loci per assay. The non-specificity and non-reproducibility of Random Amplified Polymorphic DNAs (RAPDs) markers led to the ultimate choice of present day markers such as AFLP (Amplification Fragment Length Polymorphism), STMS (Sequence Tagged Microsatellite Sites; Beckmann and Soller, 1990), SAMPL - Selectively Amplified Microsatellite Polymorphic Loci (Morgante et al., 1994) and retrotransposon based markers.

Retrotransposons are now being utilized as molecular markers in DNA finger printing, genetic linkage mapping and phylogenetic analyses. A number of features of retrotransposons make them suitable candidates for generating molecular markers in a variety of crop plants. These features of retrotransposons are as follow:

(i) They are present in high copy number in highly heterogeneous populations.
(ii) Dispersed throughout the genome.
(iii) Insertion of retrotransposon into new genomic sites occurs without losing the parental copies.
(iv) Consequences of retrotransposition range from alteration of a few hundred bases to a few kb at the site of insertion.
(v) Most retrotransposon insertions are irreversible; therefore, changes are usually fixed, which is a good attribute for phylogenetic studies (Kumar and Hirochika, 2001). Several of these elements have been sequenced and were found to display a high degree of heterogeneity and insertional polymorphism, both within and between species.

Retrotransposons consist of LTRs with a highly conserved terminus, which is exploited for primer design in the development of retrotransposon-based markers.
Figure 3B. Strategies for different types of retrotransposon based molecular Marker systems:
(a) Sequence specific amplification polymorphisms (S-SAP)
(b) Retrotransposon internal variation polymorphisms (RIVP)
(c) Inverse retrotransposon amplified polymorphisms (IRAP)
(d) Retrotransposon based insertion polymorphisms (RBIP)
Review of Literature

(Schulman et al., 2004). They have been used as DNA markers to study biodiversity in maize (Purugganan and Wessler, 1995), pea (Ellis et al., 1998), barley (Kalander et al., 1999) and pigeon pea (Maneesha, 2003) and to generate genetic linkage maps in barley, oat and pea (Kumar et al., 1997; Ellis et al., 1998; Yu and Wise, 2000). Several techniques have emerged during the last few years and are briefly outlined below (Fig. 3B):

2.10.1 S-SAP (Sequence-Specific Amplification Polymorphism)

SSAP is a multiplex amplified fragment length polymorphism (AFLP) like technique that displays individual retrotransposon insertion as bands on a sequencing gel. Fragments are amplified by PCR, using one primer from the conserved terminus of the LTR and one from the nearest restriction endonuclease site (Waugh et al. 1997a and b). A new marker technology, namely S-SAP has been employed for screening of wide range of hybrids. It is again a dominant, multiplex marker system for the detection of variation in the retrotransposon insertion site. The S-SAP strategy (Waugh et al. 1997b) that utilizes the AFLP approach to target retroelements of the genome is one of the best marker systems has recently been developed. The power of AFLP is based upon the molecular genetic variations that exist between closely related species, varieties or even cultivars. Fingerprints of a particular genotype can be generated and produced without prior sequence knowledge using a limited set of generic primers. The primers consist of a longer fixed portion (~ 15 bp) and a short (2-4 bp) random selective portion. The fixed portion gives the primer stability (hence repeatability) and the random portion allows it to detect many loci. The number of fragments detected in a single reaction can be "tuned" by selecting specific primer sets. AFLP technique is robust and reliable because of the stringent reaction conditions for primer annealing. Polymorphism is detected as presence/absence of the band (so it is usually interpreted as dominantly inherited). AFLP markers are often inherited as tightly linked clusters in centromeric and telomeric regions of
chromosomes, but randomly distributed AFLP markers also occur outside these clusters.

S-SAP involves the use of retrotransposon sequence for amplification of these regions from the genome. Fragments are amplified by PCR, using one primer designed from the conserved terminus of the LTR and one based on the presence of a nearest restriction endonuclease site. Schematic representation of the S-SAP strategy is given (Fig. 26A). Experimental procedures resemble those used for AFLP analysis. Compared to AFLP, S-SAP generally yields fewer fragments but with higher levels of polymorphism (Waugh et al., 1997b), a more even chromosomal distribution, and more co-dominance. S-SAP markers are distributed throughout the genome. It is one of the most useful marker systems in revealing a large numbers of highly polymorphic markers. The S-SAP markers require a higher initial cost of development compared to both RAPD and AFLP due to the need to isolate the LTR sequence of the retrotransposon. The S-SAP was demonstrated to be superior to both RAPD and AFLP in terms of the number of amplification products and the number of polymorphic loci. It is evident that on a per-assay basis the S-SAP procedure may be the best of the three methods for genetic analysis of the sweet potato (Berényi et al., 2002). S-SAP markers are dominant; the high multiplex ratio of the S-SAP indicates that they are more informative. An additional advantage of S-SAP system is anchoring of one primer in the nuclear genome which avoids amplification from the contaminant DNA.

It displays individual retrotransposon insertions as bands on a sequencing gel (Kumar et al., 1997, Waugh et al., 1997a and b). It is better for estimating phylogenetic relationships in plants because unlike AFLP, one of the primers is based on specific retrotransposons. This method also provides an efficient technique for evaluating retrotransposition history and behavior in natural and manipulated plant populations. It has been used to study the insertional polymorphism of the Ty1-copia group of retrotransposons in plants (Kumar et al., 1997, Waugh et al., 1997b, Ellis et al., 1999, Purugganan and Wessler, 1995,
Pearce et al., 1999, Gribbon et al., 1999, Ellis et al., 1998). S-SAP is also been used to construct linkage maps in barley, oat and pea. The heterogeneity of the internal structure within a family of PDR1 has been exploited as S-SAP markers. S-SAP markers are more informative than previously studied AFLP or RFLP markers for studying genetic diversity in the genus *Pisum* (Ellis et al., 1998). Phylogenetic trees based on the S-SAP data show that S-SAP markers for individual elements are able to resolve different species lineages within the *Pisum* genus.

### 2.10.2 Why S-SAP is most important?

The new marker technology S-SAP, which is also called retrotransposon display, has been employed for screening of wide hybrids. The dominant, multiplex marker system, S-SAP is used to detect the variations in retrotransposon flanking regions. Through S-SAP even few base pair polymorphism could be detected on PAGE. The S-SAP, a modified form of AFLP which targets the retroelements of a genome has been developed recently (Waugh et al., 1997a and b). The performance of AFLP is based upon the molecular genetic variations which exist among closely related species, varieties or even cultivars. Biodiversity assessment and phylogenetic studies are essential for preserving both land races and their related wild species, which are often disappearing rapidly (Karp et al., 1997, Kumar and Hirochika, 2001). Fingerprints of a particular genotype can be generated and produced without prior sequence knowledge using a limited set of generic primers. In S-SAP polymorphic amplification by PCR is done using one primer from the conserved terminus of the LTR and other from nearest restriction endonuclease site, generally EcoRI/MseI. Experimental procedures are similar to AFLP. Compared to AFLP, S-SAP generally yields fewer fragments but higher levels of polymorphism (Waugh et al., 1997a and b), a more even chromosomal distribution, and more co-dominance. S-SAP markers are expected to be distributed throughout the genome because genomes of plants are littered by
retrotransposons. The S-SAP displays individual retrotransposon insertions as bands on a sequencing gel (Kumar et al., 1997, Waugh et al., 1997b). It is better for estimating phylogenetic relationships in plants because unlike AFLP, one of the primers is LTR specific. This method also provides an efficient technique for evaluating retrotransposition history and behavior in natural and manipulated plant populations. It has been used to show the insertional polymorphism of the retrotransposons in various plants (Kumar et al., 1997, Waugh et al., 1997b, Ellis et al., 1999, Purugganan and Wessler, 1995, Pearce et al., 1999, Gribbon et al., 1999, Ellis et al., 1998; Pearce et al., 2000; Yu et al., 2000; Berenyi et al., 2002; Leigh et al., 2003; Galindo et al., 2004; Queen et al., 2004; Tahara et al., 2004; Madsen et al., 2005; Soleimani et al., 2005; Syed et al., 2005; Venturi et al., 2005). The heterogeneity of the internal structure within a family of PDRI has been exploited as S-SAP markers for studying genetic diversity in the genus Pisum (Vershinin et al., 1999; Jing et al., 2005). SSAP markers have been used to construct linkage maps in barley (Manninen et al., 2000; Leigh et al., 2003; Soleimani et al., 2005), oat (Yu et al., 2000), pea (Ellis et al., 1998; Pearce et al., 2000) sweet potato (Berenyi et al., 2002; Tahara et al., 2004), wheat (Queen et al., 2004), apple (Venturi et al., 2005), common bean (Galindo et al., 2004) and cashew in which it showed higher polymorphism than AFLP markers (Syed et al., 2005).

Several quantitative trait loci (QTLs) have been mapped into seven linkage groups in barley using BARE-1-SSAP analysis. In barley, retrotransposon-based monitor systems such as IRAP and REMAP in conjunction with SSAP have been used to map a locus conferring resistance to the net blotch disease caused by Pyrenophora teres on to chromosome 6H (Leigh et al., 2003; Soleimani et al., 2005). Other examples are the $M'$ locus in barley, the $Rps1k$ in soybean, syn19 and piz resistance genes to Magnaporthe grisea in pea, and the Hero locus conferring resistance to potato cyst nematode in tomato (Kumar and Hirochika, 2001, review). To study polymorphism in a large number of plants S-SAP technique has
been further modified and called “High-throughput S-SAP method”. The S-SAP’s classic approach for digestion, ligation and preamplification is combined with optimized fluorescent multiplex PCR for simultaneously selective amplifying S-SAP fragments, and multiple S-SAPs were subsequently detected by capillary electrophoresis using ABI-PRISM 3700 capillary instruments. This method is efficient, economical and accurate means for high-throughput and large-scale genotyping retrotransposon variation in plants (Tang et al., 2004)

2.10.3 IRAP (Inter-Retrotransposon Amplification Polymorphism)

This is a dominant, multiplex marker system that examines variation in retrotransposon insertion sites. IRAP fragments between two retrotransposons are generated by PCR, using outward-facing primers annealing to LTR target sequences. It does not require DNA digestion. Fragments are separated by high-resolution agarose gel-electrophoresis (Kalender et al., 1999).

2.10.4 REMAP (Retrotransposon Microsatellite Amplification Polymorphism)

REMAP fragments between retrotransposons and microsatellites are generated by PCR, using one primer based on a LTR target sequence and one based on a simple sequence repeat motif; amplification products are resolved using high-resolution agarose gel electrophoresis (Kalender et al., 1999). Several evidences suggest that repetitive sequences may have functional significance in eukaryotic genomes (Flavell, 1986). Repetitive sequences might play a role in chromosome pairing, meiotic recombination, or speciation (Flavell, 1981), however some repetitive elements appear to have no functional significance, and simply accumulate in the genome because they do not produce any phenotypic disadvantage. Microsatellites have been used for understanding genome evolution, analysis of phylogenetic relationship among related taxa, genetic mapping, DNA fingerprinting, physical mapping and gene cloning.
Simple repeat sequences are one of the major sources of molecular markers because they are also abundant and dispersed throughout the plant genome. This property has been exploited to develop a molecular marker method, REMAP. It is also a dominant, multiplex marker system that examines variation in retrotransposon insertion sites. REMAP detects polymorphism among amplicons produced between retrotransposons and microsatellites, by PCR, using one outward facing primer based on a LTR target sequence and one based on a simple sequence repeat motif (microsatellite). The fragments are separated by high-resolution agarose gel electrophoresis (Kalendar et al., 1999). REMAP has been used to distinguish between barley varieties and to produce fingerprint patterns for species across the *Hordeum* genus (Kalendar et al., 1999, 2000).

The REMAP pattern derives from short-range amplifications (hundreds of bases), so the differences showed by REMAP between individuals growing on the NFS and SFS of Evolution Canyon are likely generated by retrotransposon BARE-1 insertion, independent of other genetic changes among the individuals.

### 2.10.5 RBIP (Retrotransposon Based Insertional Polymorphism)

This is codominant marker system that uses PCR primers designed from the retrotransposon and it’s flanking DNA to examine insertional polymorphisms for individual retrotransposons. Presence or absence of insertion is investigated by two PCRs, the first using one primer from the retrotransposon and one from the flanking DNA, the second using primers designed from both flanking regions. Polymorphisms are detected by simple agarose gel-electrophoresis or by dot hybridization assays. Drawback of the method is that sequence data of the flanking regions are required for primer design. Major advantage is that RBIP does not necessarily require a gel-based detection system but can easily be adapted to automated, gel-free procedures in order to increase sample throughput (Flavell et al., 1998). RBIP is a co-dominant marker system that uses PCR primers designed from the retrotransposon and its flanking DNA sequence to examine insertional polymorphisms for individual retrotransposons. The presence or absence of
insertion is investigated by two PCRs, the first using one primer from the retrotransposon and one from the flanking DNA, the second using primers designed from both flanking regions. Polymorphism is detected by simple agarose gel-electrophoresis or by dot hybridization assays. Drawback of the method is that sequence data of the flanking regions are required for primer design. Major advantage is that RBIP does not necessarily require a gel-based detection system but can easily be adapted to automated, gel-free procedures, such as DNA chip technology in order to increase sample throughput (Flavell et al., 1998). The disadvantages of this marker system are that it requires flanking sequence information and yields co-dominant markers. It needs large investment to set them up in a species, and the chromosomal region surrounding each marker must be cloned and the host insertion site should be sequenced, the retrotransposons itself needs to be characterized before the marker development. (Flavell et al., 1998).

RBIP is less informative in one way that because only two allelic states exists (unoccupied and occupied) while other marker system occurs in multiple alleles.

Studies on biodiversity and phylogeny are critical in effectively preserving both land races and wild species of crop plants. Retrotransposon-based markers are especially suitable for studying phylogenetic relationships and genetic diversity within and between species. For instance, active retrotransposon family produces new insertions in the genome leading to polymorphism. The new insertions can be detected and used to establish the temporal sequences of insertion events, helping to determine phylogenies. These genetic properties have been exploited to study biodiversity and phylogeny in the genera *Brassica, Hordeum, Oryza* and *Pisum* (Kumar and Hirochika, 2001). A multi-retrotransposon approach has been used to estimate phylogenetic relationships between species of legumes (Pearce et al., 2000) and cereals (Gribbon et al., 1999; Kalender et al., 2000).