

***CARE1*, a *TY3-gypsy* long terminal repeat retrotransposon in the food legume chickpea (*Cicer arietinum* L)**

Manoj K. Rajput* and Kailash C. Upadhyaya

M. K. Rajput* and K. C. Upadhyaya

Molecular Genetics Laboratory,

School of Life Sciences,

Jawaharlal Nehru University,

New Delhi-110067, India

* Corresponding author

E-mail: rajputmk@hotmail.com

Abstract

We report a novel *Ty3-gypsy* long terminal repeat retrotransposon *CARE1* (*Cicer arietinum* retro-element 1) in chickpea. This 5920-bp AT-rich (63%) element carries 723-bp 5' and 897-bp 3' LTRs respectively flanking an internal region of 4300-bp. The LTRs of *CARE1* show 93.9% nucleotide identity to each other and have 4-bp (ACTA) terminal inverted repeats. A 17-bp potential tRNA^{met} primer binding site downstream to 5' LTR and a 13-bp polypurine tract upstream to 3' LTR have been identified. The order of domains (*Gag*-proteinase-reverse transcriptase-RNaseH-integrase) in the deduced amino acid sequence and phylogenetic tree constructed using reverse transcriptase sequences places *CARE1* in the *gypsy* group of retrotransposons. Homologues of a number of *cis*-elements including CCAAT, TATA and GT-1 have been detected in the regulatory region or the 5' LTR of *CARE1*. Transgenic tobacco plants containing 5' LTR:GUS construct show that its 5'-LTR is inactive in a heterologous system under normal as well as tissue culture conditions. Genomic Southern blot experiments using 5'LTR of the element as a probe show that *CARE1* or its related elements are present in the genomes of various chickpea accessions from various geographic regions.

Key words Retrotransposon *gypsy* LTR reverse transcriptase *Cicer arietinum*

Abbreviations

CARE1 *Cicer arietinum* retroelement1

LTR Long terminal repeat

RT reverse transcriptase

ORF open reading frame.

Introduction

Retrotransposons are a class of transposable elements that encode reverse transcriptase and like retroviruses propagate via a RNA intermediate but are non-infectious. These elements are widely distributed and comprise a significant fraction of eukaryotic genomes. Based on the presence or absence of LTRs retrotransposons are divided into LTR and non-LTR retrotransposons. The LTRs do not encode any known proteins, but they contain the promoters and terminators associated with the LTR retrotransposons (Kumar and Bennetzen 1999). The internal region of retrotransposons consists of *gag* (group-specific antigen) and *pol* (polyprotein) regions involved in the maintenance of the retroviral life cycle (Bishop and Varmus 1985). The *pol* gene of retrotransposon consists of four internal domains encoding the enzymes protease, reverse transcriptase, RNase H and integrase (or endonuclease). On the basis of the arrangement of the internal domains in *pol* gene, LTR retrotransposons have been classified into *Ty1-copia* and *Ty3-gypsy* groups (Doolittle et al. 1989; Xiong and Eickbush 1990). In *copia* group integrase domain is at 5' to the reverse transcriptase domain, while in *gypsy* it is located 3' to the RT domain. The terminology has been derived after the elements *gypsy* and *copia* in *Drosophila* and yeast respectively.

The great difference in genome size among the members of grass family is attributed to the presence of large amount of retroelements in plants having large

genomes (Bennetzen 2000). For example maize genome contains over 80% and barley genome more than 70% of its sequences as retroelements (Messing et al. 2004; SanMiguel and Bennetzen 1998; Vicient et al. 1999). Analysis of the recently sequenced *Arabidopsis* genome reveals that, despite its small size, it has as many as 2109 retroelements including representatives of all the categories of retroelements (Arabidopsis genome initiative 2000). Thus “C-value-paradox” i.e. non-correspondence between structural complexity to the functional complexity could largely be explained by the proportion of retroelements in the genomes. Some retrotransposons are shown to have contributed to genome evolution by changing structure, gene expression patterns and are also known to cause gene activation by inserting near or within genes or by causing alterations in transcript processing and/or stability. In some cases, insertion of retrotransposon has been shown to alter the spatial and temporal pattern of gene expression or the structure of the resulting gene product (Flavell et al. 1994; Wessler et al. 1995).

The retroelements are known to be transcriptionally and transpositionally activated in response to a variety of biotic and abiotic stresses. In that sense, they are also considered to have some role in stress alleviation phenomena in plants (Mhiri et al. 1997; Takeda et al. 1998).

Retrotransposons are now being utilized as molecular tools in DNA fingerprinting, genetic linkage mapping, phylogenetic studies and molecular breeding. The presence of retroelements in high copy number and in heterogeneous populations; their dispersion throughout the genome; and their insertion into new genomic sites without loss of parental copies are some of the properties which make them suitable candidates for generating molecular markers in various crop plants (Kumar and Bennetzen 1999).

Although legumes belong to the third largest family of angiosperms, they are least explored in terms of their genome organization (Pearce et al. 1996; Sant et al. 2000; Lall

et al. 2002; Neuman et al. 2003). Chickpea (*Cicer arietinum* L), the world's third most important food legume, is currently grown on about 10 m ha worldwide, with 95% cultivation in the developing countries like India. This self-pollinating annual diploid crop with a somatic chromosome number of $2n = 16$ has a genome size of ~931 Mb. Such a large genome is expected to have a significant fraction of retrotransposons. Here we report isolation and characterization of a typical *gypsy*-like retrotransposon *CARE1* from chickpea. The 5'-LTR of *CARE1* appears to be inactive in normal and tissue culture condition in a heterologous plant system.

Materials and Methods

Isolation of reverse transcriptase probe

DNA from chickpea (*Cicer arietinum*) variety Pusa 362 was isolated using standard procedures (Ausubel et al., 1994). Polymerase chain reaction (PCR) was performed in 50 μ l volume with 0.5 μ g of genomic DNA, each dNTP @ 200 μ M, 100 pmoles of each *gypsy* RT-specific primer 5' -TAC CCN TTN CCN CGN ATH GAY GAT-3' and 5' -GTC GTN TTY ATH GAY GAY ATH CTA-3' , 2.5 units of *Taq* DNA polymerase (New England Biolabs) and 1.7 mM MgSO₄. Temperature cycling was performed on MJ Research thermal cycler with the following profile: 94°C for 5 min, followed by 35 cycles of: 94°C for 1 min, 48°C for 1 min, 72°C for 1.5 min followed by 72°C for 10 min. The expected amplified product (270-bp) was purified from 1.2% agarose gel, cloned into pGEM-Teasy vector (Promega) and the nucleotide sequence was determined by DNA sequencer (Applied Biosystem).

Screening of chickpea genomic library

A genomic library constructed using partially digested and size-fractionated genomic DNA of chickpea, into the EcoRI site of λ ZAP vector, was screened for retrotransposon complementary sequences using the cloned RT-specific sequence as a

probe. Approximately 2 lakh plaques were plated on host *Escherichia coli* XL-1 blue, blotted onto nylon membrane and hybridized (Sambrook et al. 1989). DNA from one of the positive λ clones was isolated and the sub-fragments of retrotransposon, *CARE1* were cloned into pBluescript vector.

Sequencing of CARE1

The nucleotide sequence was determined using DNA sequencer (Applied Biosystem). The nucleotide sequence was compared with the sequences in databases using either BLASTN or BLASTX programs of NCBI (Altschul et al. 1990).

In silico analysis (Generation of dendrogram and Phylogenetic analysis)

The phylogenetic tree construction and sequence alignments using RT regions were carried out using clustal W (1.83) (<http://www2.ebi.ac.uk/clustalw>; Thompson et al., 1994). The method used is the neighbour joining method based on algorithm of Saitou and Nei (1987). The ORFs were determined by using NCBI ORF finder software (<http://www.ncbi.nlm.nih.gov/gorf/gorf.html>). The cis-acting regulatory DNA elements were scanned using PLACE database as a reference (Higo et al., 1999).

Southern analysis

Chickpea seeds (Accessions) were procured from ICRISAT (International Crop Research for Semi Arid Tropics), Patancheru, Andhra Pradesh, India. The genomic DNA digested with appropriate enzyme (Hind III) was fractionated on 0.7% agarose gel in TAE buffer and blotted onto Hybond N+ nylon membrane (Amersham). Hybridization was performed under conditions of high stringency using 5'-LTR sequence of the retrotransposon *CARE1* as a probe (Sambrook et al., 1989).

Generation of CARE1 LTR:GUS tobacco plants

The 5'-LTR region of the element was amplified using λ clone containing the full *CARE1* element as a template, 20 pmoles each of the LTR-specific primers (with added restriction sites HindIII and BamHI) 5'-AAG CTT ACT ATT GGG AAG TTG GGA

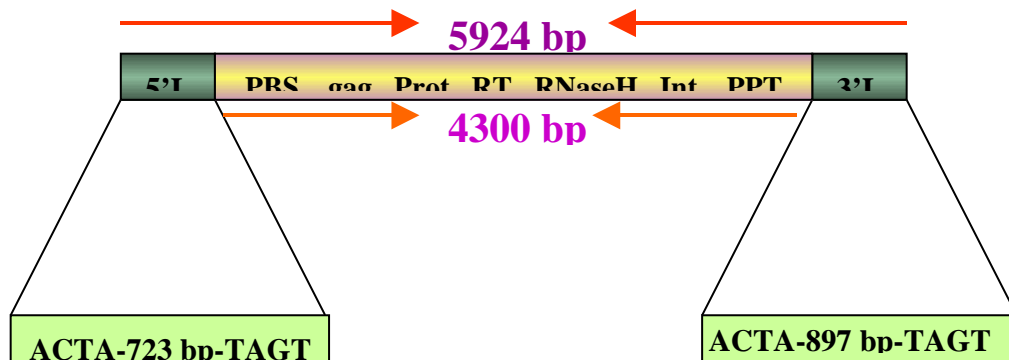
GAC T-3' and 5'-GGA TCC ACT AAA ATT GAA AAA AAG TTC C-3' in a 50 µl PCR reaction containing 100 µM dNTPs, 2 mM MgSO₄ and 2.5 units Taq DNA polymerase (New England Biolabs). Temperature cycling was performed on MJ Research thermal cycler with the following profile: 94°C for 5 min, followed by 35 cycles of: 94°C for 1 min, 53°C for 1 min, 72°C for 1.5 min followed by 72°C for 10 min. The amplified fragment (723 bp) was cloned into pGEM Teasy vector. This 723-bp fragment containing the entire 5'-LTR was excised from the plasmid and further cloned into the plant transformation vector, pBI101.

The recombinant plasmid or the construct harbouring 5'-LTR was transferred to *Agrobacterium* strain LBA4401 by freeze-thaw method. The transformed *Agrobacterium* colonies were selected on YEM-agar (yeast extract mannitol) medium supplemented with 50 µg/ml kanamycin and 25 µg/ml rifampicin. The transformed *Agrobacterium* cells were used for tobacco (*Nicotiana tabacum* cv. Xanthi) leaf-disc transformation (Horsch et al. 1985). The regeneration was carried out on MS medium supplemented with 1.0 mg/l BAP, 0.1 mg/l NAA, 250 µg/ml cefotaxime and 300 µg/ml kanamycin. After 18-25 days, well-developed independent shoots were excised and placed upright in the MS basal medium (MS medium with 0.6% agar) for rooting. The kanamycin resistant plants were tested for histochemical GUS assay. For histochemical localization of GUS, intact plants were treated with 1 mg/ml X-gluc in phosphate buffer (pH 7.2) and incubated overnight at 37°C (Jefferson 1987). The stained material was subsequently treated with 70% ethanol with several changes of solution. Southern blotting using the 5'-LTR sequence as a probe was performed to confirm the structural integrity of the construct among the individual transformants.

Results and Discussion

Isolation of retrotransposon CARE1

This is the first report of any complete retrotransposon from the food legume chickpea. A 270-bp reverse transcriptase sequence was used as a probe to screen chickpea genomic library for retrotransposon. The 5920-bp isolated retrotransposon was named CARE1 (*Cicer arietinum* Retro-element1) and submitted to database under accession no. DQ239702 (Fig. 1).



ACCGUAACACAAGTAACUA - proposed Methionyl-tRNA
PBS - TGGTATTGTGTTTCATTGAT

Fig 1. Structural features of retrotransposon CARE1 (Acc.no. DQ239702).

LTRs are shown by green rectangles at 5' and 3' ends.

Various internal domains are shown in the yellow rectangle.

The primer binding site for minus strand DNA synthesis is shown in black.

The sequence of proposed methionyl tRNA primer showing complementarity to Primer binding site is shown in light brown.

Putative polypurine tract is shown in pink.

Each LTR contains inverted terminal repeats of 4 bp (shown in enlarged boxes).

Structure of CARE1

The size of *CARE1* (5920 bp) falls in average length range of LTR-retrotransposons. The *gypsy*-like retrotransposon *Grande-1* (13.5 kb) is the largest LTR-retrotransposon, which has ever been found (Martinez-Izquierdo et al., 1997). Other *gypsy*-like retrotransposons having size 10 kb are *Athila* (10.5 kb; Pelissier et al., 1995), *Athila1.1* (12 kb; Wright and Voytas, 1998), *Cereba* (10 kb; Presting et al., 1998), *Cyclops-1* (12.3 kb; Chavanne et al., 1998) and *RIRE3* (10.5 kb; Kumekawa et al., 1999b). The AT rich (63%) element *CARE1* carries 723-bp and 897-bp 5' and 3' LTRs respectively flanking an internal region of 4300 bp with two successive domains, both showing homology to retroviral genes *gag* and *pol*. The *pol* region of *CARE1* contains sequence motifs related to the enzymes protease (PR), reverse transcriptase (RT), RNase H (RH) and integrase (IN) in the order 5'-PR-RT-RH-IN-3' characteristic for a *gypsy*-like retrotransposons and retroviruses. Like certain retroelements such as marY1 and Tnt1, *CARE1* also lack typical canonical sequence 5'-TG.....CA-3' (Grandbastien et al., 1989; Murata and Yamada, 2000). Since a typical canonical sequence is not present in *CARE1* LTRs, it is probable that (CA) might have mutated to (GA). Both LTRs of *CARE1* have perfect 4-bp (ACTA) inverted repeats (IRs), which are thought to be important for integration or insertion. In *CARE1* we identified a potential 17-bp (TGGTATTGTGTTTCATTG) non-overlapping tRNA^{met} primer binding (site for minus strand DNA synthesis which is located just down stream to the 5'LTR and a 13-bp (AAAAGAGGAGAAA) non-overlapping polypurine tract just upstream to the 3'LTR, which could be primer binding site for plus-strand DNA synthesis. In most of retroelements including *CARE1* a 16-21 nucleotide stretch of PBS sequence begins with TGG and shows best matching to cytosolic tRNA^{met} of plants suggesting, that cytosolic tRNA^{met} from chickpea is used to prime the minus-strand DNA synthesis. The other *gypsy*-like retroelements with non-overlapping PBS and PPT are

412, *Micropia*, *Ulysses*, *SURL*, *Mag*, *Del* and *Ty3*. The PBS of *CARE1* also shows significant homologies to the PBS of *gypsy*-like retrotransposons *LORE1*, *Del-1*, *Magellan* (Madsen et al., 2005; Smyth et al., 1989; Purugganan et al., 1994) and most other LTR-retrotransposons. Reverse transcription of retroelement transcripts has always been found to proceed from a tRNA primer bound at the 3' end of the 5'LTR. For the LTR-retrotransposons, the process of reverse transcription can be conveniently separated into minus-strand and plus-strand reverse transcription. In the former, the mRNA of the retrotransposon serves as a template for reverse transcription but, in the latter the minus-strand reverse transcript (cDNA) serves as template (Boeke and Cores, 1989).

Structure of long terminal repeats (LTRs) of CARE1

The 3' -LTR is longer by 174 bp. This 174-bp stretch contains 165 duplicated segment and two additional sequences of 5 and 4 bp. The 5-bp (GTAAT) sequence in 3' -LTR is located as an intervening sequence between the 165-bp duplicated segment, and the 4-bp (TCAG) is located 9-bp upstream to the end of the 3' -LTR (**Fig. 2a & 2b**). Excluding the duplicated segment, the two LTRs show a homology of ~94% at the nucleotide level. One deletion of a single nucleotide (indel) is present in both the LTR's. The 165-bp segment that is duplicated in the 3' -LTR shows a homology of 87.8% with the corresponding region in the 5' -LTR. The duplicated region shows ~90% homology and contains 12 nucleotide substitutions, which includes 7 transversions (4 T to G, one G to T, one G to C and one T to A) and 5 transitions (2 T to C, 2 C to T and one G to A). The 12 nucleotide changes possibly indicate that the duplication event may not be a recent one.

```

5' LTR          ACTATTGGGAAGTTGGGAGACTTGTGCCCTAACTGAG 37
3' LTR          ACTATTGGGAAATTTGGGAGACTTGTCCCTAGTAGAG 37
***** ** ***** **

```

5' LTR AACTCAAATTCGACCGTTTCTAACACCATTGAAAGCAGTTTAAACAAGAATCATTCATGA 97
3' LTR AACTCAAAGACAACAATTTCTAACACCATTTGAGCAGTTTATCAAGAATAACTCATGA 97
***** * ** ***** * *****

5' LTR ATCTTTTTTTTAAATCTTCTCTAATCTCTATCTTTTCTATCTATGTCATGAGTAACTAAA 157
3' LTR ATCCTTCTTGTAATCTTCTCTAATCGCTATCTTTTCTATCTCTATCATGAGTAACTAAA 157
*** ** * ***** * *****

5' LTR CCCTATTTGTTAGGAATGAGTGTAACAACTAAAACCCTTATTTTTATGATTTGATTTCT 217
3' LTR CCCTATTTGTAAGGGAGGAGTGAATAAGATGAAACCCTTATTTTTCTGATTTGATTTCT 217
***** ** * ***** ** * *****

5' LTR AGCTATATGAATGAGTTTATTGAATTATTTTTCTCATCTCTGTGCTTAATGCTTATTATT 277
3' LTR AGTTATATGAATGAGTTTATTAAATTATTTTTCTCATCTCTGTGCTTAATACTTTTATT 277
** ***** ***** *****

5' LTR GCTTGA----- 283
3' LTR GCTTGAGTAATTAATCTCTATCTTTTCTATCTATCATGAGTAATTAACCCTATTTTT 337

<----G-----duplication-----G-->

5' LTR ----- 283
3' LTR TATGGATGAGCGTAATAAGATGAACTCTTATTTCTCTGATTTGATTTATAGTTATATGA 397
A-G---G---T-----C-----T-----C-----

5' LTR -----TCAA 287
3' LTR ATGAGTTTATTGAATTATTTTTCTCATCTTGGTGCTTAATACTTTTATTGCTTGATCAA 457

-----A-----CT----->

5' LTR CATTAAA-TGTTCTACGATTCGTATTTTGAATGGGAGTGGACTTTACGAATGCTTGAGA 346
3' LTR CATTAAAATGTTCTACAATTCGTATTTTGAACGGAAGTGGACTTTACGAATGCTTGAGA 517
***** ***** ***** * *****

5' LTR TGAGAAATTCATGAATTTTAACTATGGATAGATACATGTCATGAAACTAATTAATTA 406
3' LTR TGAGAAATTCATGATATTGTAGTCTAGGGATAGATGCAGGTCGTAACCAATCAAATTA 577
***** ** * ***** * *****

5' LTR GTTGCAAAGGAAATAGTTTAAACAAGATAACTCTTATACAATAATGCTTAATTTCTAATCTT 466
3' LTR CTTGCAAATGCAATAGTTTAAACAAGAGAATTCATGTACGTTATGGCTTAATTTCTAATCTT 637
***** * ***** * *****

5' LTR AAATCCATTAAGGAATTAGGGTTACTTTGGAATTAAGGTTTGTCACTAAGACATTAG 526
3' LTR AAATCCACTAAGGAATTATGGTTACTTTGGAATTAAGGTTTGTCACTAAGACATTAG 697
***** ***** *****

5' LTR GGCTAGAATAATAAAGAGAATTCGATAATAATCAATAAAGGAATTTAATACTAGGATC 586
3' LTR GGAAAAGAATAATAAAGAGAATTCGATAATAATCAATAAAGAAATTCATAACTAGGATC 757
** ***** *****

5' LTR AAATTAGACACCAAGGTTGGATTGGAAGTGAAGTCAATCCCTGACATTTTTCTTATTAT 646
3' LTR AAATTAGAAATTATGGTTGAATTCGAAGTGAAG-CTCATCCCTGACATTTTTCTTATTAT 816
***** * * ***** *****

5' LTR AAAGGATCAATTTTAGTACTGTTGTTAATTTTAAATATTATTCAATTAATTTGGGAACT 706
3' LTR AAAGGATCAATTTTACTACCGTTGTTAATTTTAAATATTGTTTCAATCAATTTCTGAACT 876
***** ***** ***** ***** * *****

5' LTR TTTTGTTT-----AATTTTAGT 723
3' LTR TCTCGTTTTTCAGGATTTAGT 897
* * *** * *****

Figure 2a. Alignment of nucleotide sequences of 5'LTR and 3'LTR.

The stars denote identical bases. In red duplication of internal sequence of 3'LTR is shown (bases in red are those which are different in duplicated sequence).

```

5' LTR                                TAATCTCTATCTTTTCTATCTATGTCATGAGTAACTAAA 157
3' LTR                                TAATCGCTATCTTTTCTATCTCTATCATGAGTAATTAAA 157
3' LTR duplication >                 289 TAATCTCTATCTTTTCTATCTCTATCATGAGTAATTAAA 328
                                     ***** * ***** * ***** * *****
                                     *
                                     *
                                     *
                                     *
                                     *
5' LTR                                CCCTATTTGTTAGGAATGAGTGTAAACAACTAAAACCCTTATTTTATGATTGATTTCT 217
3' LTR                                CCCTATTTGTAAGGGAGGAGTGTAAATAAGATGAAACCCTTATTTTCTGATTGATTTCT 217
3' LTR dup.                          CCCTATTTTATGGATGAGCGTAATAAGATGAAACTCTTATTTCTCTGATTGATTTAT 388
                                     ***** * * * * * * * * * * * * * * * * * * * * * * * * * * * * *
                                     * *
                                     *
                                     *
                                     *
                                     *
5' LTR                                AGCTATATGAATGAGTTTATTGAATTATTTTTCTCATCTCTGTGCTTAATGCTTATTATT 277
3' LTR                                AGTTATATGAATGAGTTTATTAAATTATTTTTCTCATCTCTGTGCTTAATACTTTTATT 277
3' LTR dup.                          AGTTATATGAATGAGTTTATTGAATTATTTTTCTCATCTTGGTGCTTAATACTTTTATT 448
                                     ** ***** * * * * * * * * * * * * * * * * * * * * * * * * * * *
                                     *
                                     *
                                     *
                                     *
                                     *
5' LTR                                GCTTGA 283
3' LTR                                GCTTGA 283
3' LTR dup.                          GCTTGA 454
                                     *****

```

Figure 12b. Alignments of duplicated sequence with its original in 3'LTR and homologous sequence in 5'LTR.

Black * shows bases matching in all three sequences.
 Red * shows bases matching in 5'LTR and duplicated 3'LTR.
 Blue* shows bases matching in 5'LTR and original 3'LTR.
 shows bases matching in original 3'LTR and duplicated 3'LTR.

The potential functionality of a retrotransposon depends on its regulatory sequences (promoter) present in its 5' -LTR; however, the transposability of a retrotransposon depends on the functionality of the translated polyprotein as well as on the activity of regulatory sequences or *cis*-elements. While scanning the 5' -LTR of *CARE1* for *cis*-elements using PLACE (Plant *cis*-acting elements) database as a reference (Higo et al., 1999) three putative TATA boxes, CCAAT box, GT-1 consensus, copper

responsive element GATC, ethylene responsive element ERELEE4 and various MYB related *cis*-elements were detected. The 5' -LTR contains homologues of MYB related elements which are shown to be induced by wounding and elicitors in tobacco retrotransposon *Tto1* and defense-related genes (Sugimoto et al., 2000), ethylene responsive element ERELEE4 of retrotransposon *TLC1.1* of *Lycopersicon chilense* (Tapia et al., 2005), a copper responsive element GATC (Quinn and Merchant, 1995).

When 5' -LTR of *CARE1* is compared with 5' -LTRs of active retrotransposons *LORE1*, *Tnt1* and *Tto1*, it is found that LTRs of *CARE1* contain almost all *cis*-elements which are present in their LTRs (Grandbastien et al., 1994; Sugimoto et al., 2000). It may be possible that some of these *cis*-elements could be binding sites for repressors, which after binding may stop the recruitment of transcription factors to the promoter.

Although most gypsy-like retrotransposons show no divergence between their two LTRs, such sequence differences have been observed in many gypsy-like elements (Smyth et al., 1989; Chavanne et al., 1998). The 723-bp 5' -LTR and 897-bp 3' -LTR of *CARE1* with 93.9% inter-LTR sequence identity fall within the average size range of LTRs of most of the gypsy-like retrotransposons (Smyth et al., 1989; Pelissier et al., 1996; Chavanne et al., 1998; Wright and Voytas, 1998; Kumekawa et al., 1999). This suggests that LTRs may diversify very frequently, presumably because LTRs are non-coding regions. The analyses of the LTRs of *CARE1*, *Del-1*, *Micropia*, *Zeon-1*, *Cyclops-2*, *Tnt1*, *Tto1*, *Tos17* and *LORE1* indicate that there appears to be a relation between functionality and inter-LTR homology for a given retroelement (Grandbastien et al., 1989; Smyth et al., 1989; Purugganan et al., 1994; Grandbastien et al., 1994; Hirochika et al., 1996; Kumar et al., 1999; Madsen et al., 2005). To date only a few active retrotransposons are known, three of them *Tnt1* (tobacco), *Tto1* (tobacco), *Tos17* (rice) are *copia*-like (Grandbastien et al., 1994; Lucas et al., 1995; Hirochika et al., 1996; Mhiri et al., 1997; Takeda et al., 1998; Kumar et al., 1999) but, recently discovered active element *LORE1* (*Lotus japonicus*) belongs to the gypsy group (Madsen et al., 2005). The homology between 5' -LTR and 3' -LTR of these active retroelements is very high.

Internal structure of CARE1

The deduced amino acid sequence of the internal region of *CARE1* shows the presence of five open reading frames along the same strand and in the same frame. The inter-ORF regions contain many termination (stop) codons. All conserved domains (*gag*, proteinase, reverse transcriptase and RNase H) except integrase were detected in this +1 frame. The integrase domain was found in deduced amino acid sequence in +2 frame. The first ORF is predicted to encode a protein containing 109 amino acids with a putative central motif QGX₂EX₅FX₃LX₂H, from amino acid 109 to 126, which is common to retroviridae *gag* proteins, but is poorly conserved. The predicted molecular mass of this

protein is very much less than those of most *gag* gene products, which is generally ~60-80 kDa. In *CARE1*, a proteinase conserved domain LIDIGA is present from 428 to 433 amino acid positions. The aspartic proteinases of retroelements are extremely diverse and hence, little sequence conservation outside the active site is observed between *CARE1* and others (Smyth et al., 1989; Purugganan et al., 1994; Chavanne et al., 1998; Kumekawa et al., 1999).

The second or the smallest ORF extends from 2545 to 2652 nucleotide positions and is followed by reverse transcriptase region with conserved domain from 2998 to 3493 nucleotide positions. The alignment of RTase conserved region of *CARE1* with the corresponding regions of retroelements from various organisms (plants, animals, viruses, fungi and insects) reveals highly conserved motifs like PRID and FXXD. These regions are highly conserved and they have been used to design degenerate primers to amplify RTases of *gypsy*-like retrotransposons from diverse plant species (Muthukumar and Bennetzen, 2004). In *CARE1*, the actual conserved motif is larger which contains the core motif FIND. The third or the longest ORF starts at the 3523rd nucleotide and ends at 3906; and it is predicted to encode polypeptide of 128 amino acids with three internal methionines. This ORF is immediately followed by fourth ORF, which starts at nucleotide position, 3919 and ends at 4074. RNase conserved domain is present in this ORF from 1036 to 1039 amino acid positions. For retroviruses, an RNaseH conserved motif is represented by TDGS (thr, asp, gly, and ser) while it is TDAS (thr, asp, ala and ser) for most *gypsy*-like elements (Warmington et al., 1985; Smyth et al., 1989) but in case of *CARE1*, *Cyclops-2* and *Ulysses* it is CDAS (cys, asp, ala and ser). The 189-bp fifth ORF exists between 4075 and 4263 nucleotides. After fifth ORF, a single nucleotide frame shift occurs which separates the integrase region from the whole element in the context of translational frame and the integrase conserved domain is present between 4259 to 4744 nucleotides and zinc finger motif in this region is from 1271 to 1292 amino

acid positions. This deduced amino acid sequence of *CARE1*, which corresponds to the endonuclease region of other retrotransposon, shows an amino acid configuration capable of forming a zinc-binding domain, which could interact with DNA to facilitate integration of LTR-retroelements into the host genome (Grantgenett and Mumm, 1990). The structure of conserved zinc finger motif in *CARE1* is His-(7)-Lys-(10)-Cys-(1)-Cys. Although this conserved motif of *CARE1* shows high homology to the corresponding regions of other retrotransposons but does not represent a typical structure. Functional polyproteins are known to form in retroelements by frame-shifting (Farabaugh et al., 1993). However, due to presence of many chain terminating codons in *CARE1*, it appears that this copy of the element is a non-functional one but a possibility of another functional element of the *CARE* family can not be ruled out.

Phylogenetic Analysis

The phylogenetic analysis of genomic sequences is done on the basis of their deduced amino acid homology. Among retroelements from the diverse organisms the region encoding the reverse transcriptase is highly conserved (Feng and Doolittle, 1987; Xiong and Eickbush, 1990; McClure, 1991), thus these are used to derive the phylogenetic tree along with RTase of *CARE1* by neighbor joining method (Saitou and Nei, 1987). All categories of LTR retroelements viz. *copia*, *gypsy* and retroviruses were used for alignment and generation of phylogenetic trees (**Table 1**).

Table: 1. Retroelements used for alignments

Accession no.	Organism	Category
D83003	<i>Nicotiana tabacum</i>	<i>copia</i>
X13777	<i>Nicotiana tabacum</i>	<i>copia</i>
AC087545	<i>Oryza sativa</i>	<i>copia</i>
DQ105074	<i>Malus sp</i>	<i>copia</i>
AY654426	<i>Citrus sp</i>	<i>copia</i>
AF416818	<i>Musa sp</i>	<i>copia</i>
AJ414059	<i>Brassica sp</i>	<i>copia</i>

DQ100159	<i>Camellia sp</i>	<i>copia</i>
DQ054417	<i>Pinus</i>	<i>copia</i>
DQ54416	<i>Ephedra</i>	<i>copia</i>
AF295692	<i>Anopheles</i>	<i>copia</i>
AF398212	<i>Zea mays</i>	<i>copia</i>
AF130856	<i>Phytophthora sp</i>	<i>copia</i>
AF231939	Sweet potato	<i>copia</i>
M94164	<i>S. cerevisiae</i>	<i>copia</i>
M94494	<i>H. annus</i>	<i>copia</i>
AJ966990	<i>Lotus japonicus</i>	<i>gypsy</i>
AJ000640	<i>Pisum sativum</i>	<i>gypsy</i>
AJ290632	<i>Pinus sp</i>	<i>gypsy</i>
AJ971814	<i>Musa sp</i>	<i>gypsy</i>
AJ415653	<i>Brassica sp</i>	<i>gypsy</i>
AY959279	<i>Cycas sp</i>	<i>gypsy</i>
AY683037	mungbean	<i>gypsy</i>
AJ295135	<i>Pisum sativum</i>	<i>gypsy</i>
AJ295132	<i>Zea mays</i>	<i>gypsy</i>
AJ295139	<i>Arabidopsis</i>	<i>gypsy</i>
AJ243725	Frog	<i>gypsy</i>
AB033239	<i>Chlamydomonas</i>	<i>gypsy</i>
X03734	<i>Drosophila</i>	<i>gypsy</i>
AY877248	HIV-1virus	Retrovirus
AY454232	lentivirus	Retrovirus
AF126467	Simianretrovirus	Retrovirus
U34348	Humanendovirus	Retrovirus
M25768	Humanrt	Retrovirus

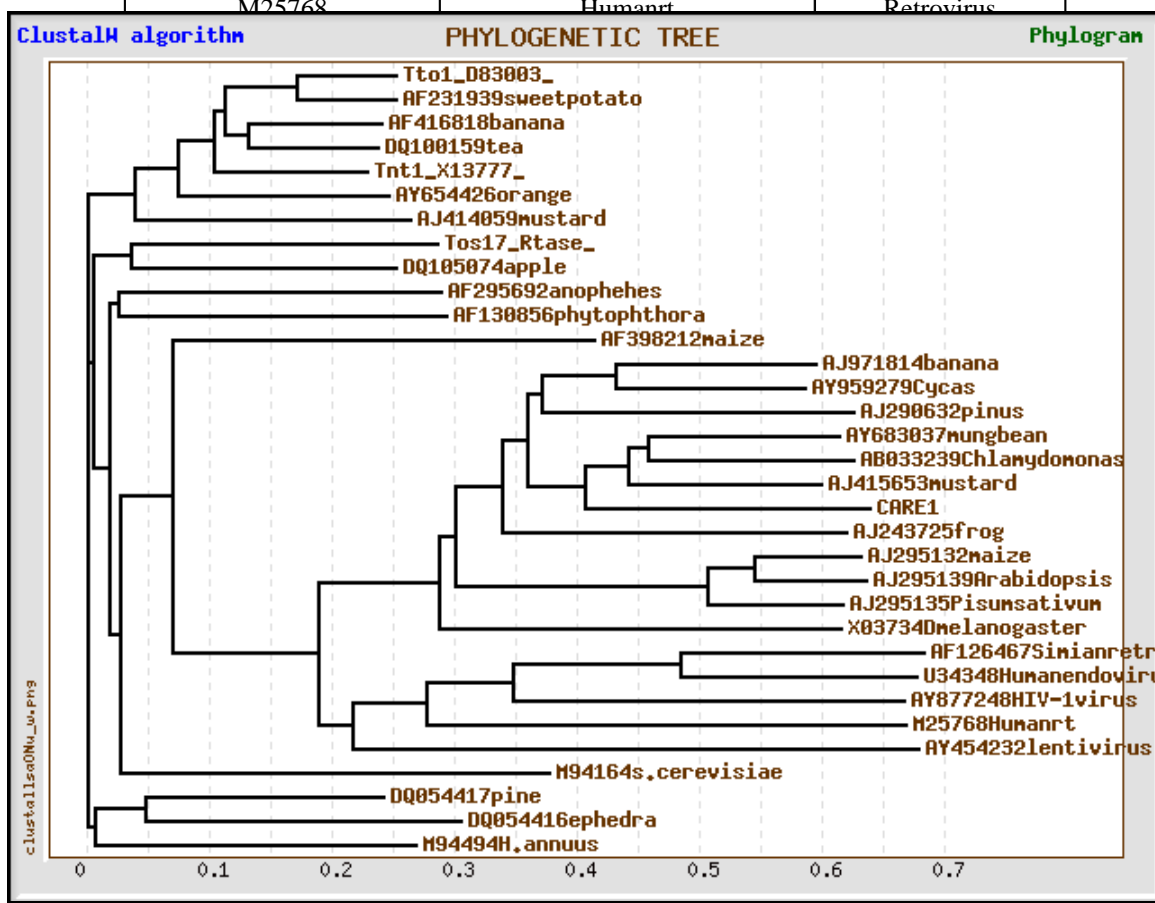


Figure 3. Phylogenetic tree of predicted amino acid sequences of RTase of CARE1 and other diverse retroelements. See Table 4 for source of these elements. Divergences in distance units are indicated by branch lengths.

In the phylogenetic tree *CARE1* has been shown to branch from a point, which is common to other *gypsy*-like retrotransposons from diverse organisms (**Fig. 3**). Also the *gypsy*-like retrotransposons and retroviruses are observed to have common origin in the phylogenetic tree. This shows that retroviruses and *gypsy*-like retrotransposons diversified from a common ancestor. Alignments reveal that, *CARE1* reverse transcriptase shows 26% to 59% amino acid sequence identity to the RT regions of other *gypsy*-like elements and 2-12% to the RT regions of *copia*-like elements and retroviruses.

The LTRs of a retroelement are very specific to the element and are quite conserved among their families but the conservation of other regions like *gag* and RTase is not restricted to a particular family. So the distribution, organization and presence of *CARE1* family representatives in various chickpea accessions were assessed by Southern blotting using the 5' -LTR of *CARE1* as a probe. The results indicated that not only populations of *copia*-like retrotransposon sequences (Sant et al., 2000) are accumulated in chickpea genome but a sizable fraction of chickpea genome is also contributed by *gypsy*-like sequences.

Functional analysis of CARE1 promoter (5' -LTR) in heterologous plant

Considering a number of stop codons in the reading frames of the CARE1 retrotransposon, it appears that the isolated copy of the element is possibly non-functional. However, it is possible that the 5' -LTR, that contain the regulatory sequence may be functional. In order to determine the functionality of 5' -LTR, a 5' -LTR-GUS construct was made and transferred to tobacco. The structural integrity of the construct in transgenic tobacco plants were assessed by Southern blotting. The GUS expression was monitored in the transgenic plants. However, no GUS expression either histochemically or spectrofluorometrically could be detected in the normal or in response to stress. Non-expression of GUS indicates that the 5' -LTR is inactive as a promoter. This inactivity could be due to defective 5' -LTR or additional sequences are required for its functionality.

The cloning and characterization of the retrotransposon *CARE1* from the legume chickpea is the first step towards the study of retrotransposons in this crop. Future studies would focus on the suitability and applicability of this retrotransposon in generating molecular markers for phylogenetic relationships among various chickpea varieties.

Acknowledgements

The authors would like to extend their sincere thanks to Dr. KR Koundal, National Research Centre on Plant Biotechnology, IARI, New Delhi, for chickpea genomic library and UGC, govt. of India for fundings. Dr. M K Rajput thanks to Council of Scientific and Industrial Research, govt. of India for fellowship.

References

Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* 25:3389-3402

Ausubel MF, Brent R, Kungston R, Moore D, Seidman J, Smith J, Struhl K (1994) Current Protocols in Molecular Biology. New York: John Wiley and Sons, Inc, pp 2.3.3-2.3.5

Bennetzen JL (2000) Transposable element contributions to plant gene and genome evolution. *Plant Mol. Biol* 42:251–269

Chavanne F, Zhang D, Liaud M, Cerff R (1998) Structure and evolution of *Cyclops*: a novel giant retrotransposon of the *Ty3/Gypsy* family highly amplified in pea and other legume species. *Plant Mol Biol* 37:363-375

Doolittle RF, Feng DF, Johnson MS, McClure MA (1989) Origins and evolutionary relationships of retroviruses. *Quat Rev Biol* 64:1-30

[Farabaugh PJ](#), [Zhao H](#), [Vimaladithan A](#) (1993) A novel programmed frameshift expresses the POL3 gene of retrotransposon Ty3 of yeast: frameshifting without tRNA slippage. *Cell* 74:93-103

Feng DF, Doolittle RF (1987) Progressive sequence alignment as a prerequisite to correct phylogenetic tree. *J Mol Evol* 25:351-360

Feschotte C, Jiang N, Wessler SR (2002) Plant transposable elements: where genetics meets genomics. *Nat Rev Genet* 3:329-341

Flavell AJ, Pearce SR, Kumar A (1994) Plant transposable elements and the genome. *Curr Opin Genet Dev* 4:838-844

Grandbastien MA, Spielman A, Caboche M (1989) *Tnt1*, a mobile retroviral-like transposable element of tobacco isolated by plant cell genetics. *Nature* 337:376-380

Grandbastien MA, Audeon C, Casacuberta JM, Grappin P, Lucas H, Moreau C, Pouteau S (1994) Functional analysis of the tobacco *Tnt1* retrotransposon. *Genetica* 93:181-189

Grandgenett DP, Mumm SR (1990) Unraveling retrovirus integration. *Cell* 60:3-4

Higo K, Ugawa Y, Iwamoto M, Korenaga T (1999) Plant cis-acting regulatory DNA elements (PLACE) database: 1999. *Nucleic Acids Res* 27:297-300

- Hirochika H, Otsuki H, Yoshikawa M, Otsuki Y, Sugimoto K, Takeda S (1996) Autonomous transposition of the tobacco retrotransposon *Tto1* in rice. *Plant Cell* 8:725-734
- Horsch RB, Rogers SG, Fraley RT (1985) Transgenic plants. *Cold Spring Harb Symp Quant Biol* 50:433-437
- Jefferson RA, Kavanagh TA, Bevan MW (1987) GUS fusions: beta-glucuronidase as a sensitive and versatile gene fusion marker in higher plants. *EMBO J* 6:3901-3907
- Kumar A, Bennetzen JL (1999) Plant retrotransposons. *Annu Rev Genet* 33:479-532
- Kumekawa N, Ohtsubo H, Horiuchi, Ohtsubo E (1999) Identification and characterization of novel retrotransposons of the *gypsy* type in rice. *Mol Gen Genomics* 260:593-602
- Lall IP, Maneesha, Upadhyaya KC (2002) *Panzee*, a *copia*-like retrotransposon from the grain legume, pigeonpea (*Cajanus cajan* L.). *Mol Genet Genomics* 267:271-280.
- Lucas H, Feuerbach F, Kunert K, Grandbastien MA, Caboche M (1995) RNA-mediated transposition of the tobacco retrotransposon *Tnt1* in *Arabidopsis thaliana*. *EMBO J* 14:2364-2373
- Madsen LH, Fukai E, Radutoiu S, Yost CK, Sandal N, Schauser L, Stougaard J (2005) *LORE1*, an active low-copy number *Ty3-gypsy* retrotransposon family in the model legume *Lotus japonicus*. *Plant J* 44:372-381
- Manninen I, Schulman AH (1993) *BARE-1*, a *copia*-like retroelement in barley (*Hordeum vulgare* L.). *Plant Mol Biol* 22: 829-846
- McClure MA (1991) Evolution of retrotransposons by acquisition or deletion of retrovirus-like genes. *Mol Biol Evol* 8:835-856
- Messing J, Bharti AK, Karlowski WM, Gundlach H, Kim HR, Yu Y, Wei F, Fuks G, Soderlund CA, Mayer KF, Wing RA (2004) Sequence composition and genome organization of maize. *Proc Nat Acad Sci USA* 101:14349-14354
- Mhiri C, Morel JB, Vernhettes S, Casacuberta JM, Lucas M, Grandbastien MA (1997) The promoter of the tobacco *Tnt1* retrotransposon is induced by wounding and by abiotic stress. *Plant Mol Biol* 33:257-266

Muthukumar B, Bennetzen JL (2004) Isolation and characterization of genomic and transcribed retrotransposon sequences from sorghum. *Mol Gen Genomics* 271:308-316

Neumann P, Pozarkova D, Macas J (2003) Highly abundant pea LTR retrotransposon *Ogre* is constitutively transcribed and partially spliced. *Plant Mol Biol* 53:399-410

Pearce SR, Harrison G, Li D, Heslop Harrison JS, Kumar A, Flavell AJ (1996) *Ty1-copia* group retrotransposons in *Vicia sps*: copy number, sequence heterogeneity and chromosomal location. *Mol Gen Genet* 250:305-315

Pelissier T, Tutois S, Tourmente S, Deragon JM, Picard G (1996) DNA regions flanking the major *Arabidopsis thaliana* satellite are principally enriched in *Athila* retroelement sequences. *Genetica* 97:141-151

Purugganan MD, Wessler SR (1994) Molecular evolution of *Magellan*, a maize *Ty3/gypsy* like retrotransposon. *Proc Nat Acad Sci USA* 91:11674-11678

Quinn JM, Merchant S (1995) Two copper-responsive elements associated with the *Chlamydomonas* *Cyc6* gene function as targets for transcriptional activators. *Plant Cell* 7:623-628

Saitou N, Nei M (1987) The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol Biol Evol* 4:406-425

Sambrook J, Fritsch EF, Maniatis T (1989) *Molecular Cloning. A Laboratory Manual*, 2nd edn. Cold Spring Harbor: Cold Spring Harbor Laboratory Press

SanMiguel P, Gaut BS, Tikhonov A, Nakajima Y, Bennetzen JL (1998) The paleontology of intergene retrotransposons of maize. *Nat Genet* 20:43-45

Sant VJ, Sainani MN, Sami-Subbu R, Ranjekar PK, Gupta VS (2000) *Ty1-copia* retrotransposon-like elements in chickpea genome: their identification, distribution and use for diversity analysis. *Gene* 257:157-166

Smyth DR, Kalitsis P, Joseph JL, Sentry JW (1989) Plant retrotransposon from *Lilium henryi* is related to *Ty3* of yeast and the *gypsy* group of *Drosophila*. *Proc Nat Acad Sci USA* 86:5015-5019

Sugimoto K, Takeda S, Hirochika H (2000) MYB-Related transcription factor NtMYB2 induced by wounding and elicitors is a regulator of the tobacco retrotransposon *Tto1* and defense-related genes. *Plant Cell* 12:2511-2527

Takeda S, Sugimoto K, Otsuki H, Hirochika H (1998) Transcriptional activation of the tobacco retrotransposon *Tto1* by wounding and methyl jasmonate. *Plant Mol Biol* 36:365–376

Tapia G, Verdugo I, Yanez M, Ahumada I, Theoduloz C, Cordero C, Poblete F, Gonzalez E, Ruiz-Lara S (2005) Involvement of ethylene in stress-induced expression of the TLC1.1 retrotransposon from *Lycopersicon chilense* Dun. *Plant Physiol* 138:2075-2086

The Arabidopsis Genome Initiative (2000) Analysis of genome sequence of the flowering plant *Arabidopsis thaliana*. *Nature* 408:796-815

The International rice genome sequencing project (2005) The map-based sequence of the rice genome. *Nature* 436:793-800.

Thompson JD, Higgins DG, Gibson TJ (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res* 22:4673-4680

Vicient CM, Kalendar R, Anamthawat-Jonsson K, Schulman AH (1999) Structure, functionality, and evolution of the *BARE-1* retrotransposon of barley. *Genetica* 107:53-63

Warmington JR, Waring RB, Newlon CS, Indge KJ, Oliver SG (1985) Nucleotide sequence characterization of *Ty* 1-17, a class II transposon from yeast. *Nucleic Acids Res* 13:6679-6693

Wessler SR, Bureau TE, White SE (1995) LTR-retrotransposons and MITEs: important players in the evolution of plant genomes. *Curr Opin Genet Dev* 5:814-821

Wright DA, Voytas DF (1998) Potential retroviruses in plants: *Tat1* is related to a group of *Arabidopsis thaliana* *Ty3/gypsy* retrotransposons that encode envelope-like proteins. *Genetics* 149:703-715

Xiong Y, Eickbush TH (1990) Origin and evolution of retroelements based upon their reverse transcriptase sequences. *EMBO J* 9:3353-3360

