Isolation and Characterization of retroelements from Chickpea (*Cicer arietinum*) genome

Dissertation Submitted to the
Jawaharlal Nehru University
For the award of the degree of

MASTER OF PHILOSOPHY

MANOJ KUMAR

School of Life Sciences
Jawaharlal Nehru University
New Delhi - 110067
INDIA

2003

SCHOOL OF LIFE SCIENCES
JAWAHARLAL NEHRU UNIVERSITY
NEW DELHI - 110067

CERTIFICATE

The research work embodied in this dissertation entitled **'Isolation and Characterization of retroelements from chickpea (*Cicer arietinum*) genome'** has been carried out at the School of Life Sciences, Jawaharlal Nehru University, New Delhi. This work is original and has not been submitted so far in part or in full, for the award of any degree or diploma of any university.

Manoj Kumar
(Candidate)

Prof. K. C. Upadhyaya
(Supervisor)

Prof. A. Bhattacharya
(Dean)

July 2003
School of Life Sciences
Jawaharlal Nehru University
New Delhi - 110067
INDIA

Dedicated to Mr. John. G. Mendel,
the father of Genetics

Acknowledgements

Acknowledgements

With ineffable gratitude, I sincerely acknowledge my supervisor Prof. K. C. Upadhyaya for invaluable guidance and encouragement he rendered me from time to time. With deep sense of reverence I thank him for suggesting me the area of research and for promptly providing the requisite facilities for smooth pursuance of my work. I am grateful to him for giving full freedom at work.

I acknowledge Prof. A. Bhattacharya, the present Dean and Prof. R. Saxena, former Dean, School of Life Sciences, for extending all necessary facilities during the course of work.

I am heartened to acknowledge my senior colleagues in the lab I.P. Singh, Ritu, Kavita, Amita, Maneesha, Anita, Rashmi and Purnima for their unforgettable company in the lab.

I feel previliged to have friends like Surendra, Jitendra, Hero, Kp, Kp, Sanjay, Angna, Rahul, Vikas, Kamal, Sanghmitra, Ganesha, Zubaida, Dola, Vijay and Bimla who always cheered me and were always available whenever needed.

I acknowledge the services of Mishra ji, sh. Ajit Singh and Anwar in the lab.

The administrative support of Mr. K. Raghunathan, AO, madam Meenu and other staff of SLS, CPMB staff member Mr. M. M. Sharma; and the technical help of CIF staff: Mr. A. C. Alexander, Mr. B. A. Khan and Mr. S. K. Mishra is duly acknowledged. I extend my sincere thanks to Dr. Andrew Lynn, faculty BIC and P.V.S for their suggestion during computer analysis.

I thank all members of JNU and NCPGR for providing necessary services.

Financial assistance from CSIR in the form of junior research fellowship is greatly acknowledged.

Last but not least I would like to thank my mother, my brother and my sister, without their support, encouragement and moral support the work would not have come out in this form

(Manoj Kumar Rajput)

Abbreviations

A	Adenosine
a.a	amino acid
AFLP	Amplified Fragment Length Polymorphism
APS	Ammonium per sulphate
ATP	adenosine tri-phosphate
bp	base pair(s)
C	Cytosine
^0C	Degree Celsius
cpm	Counts per minute
CTAB	Hexa decyl trimethyl ammonium bromide
dATP	deoxy adenosine tri-phospahte
dCTP	deoxy cytosine triphosphate
dGTP	deoxy guanine triphosphate
ddNTP	Dideoxy nucleotide triphosphate
DTT	Dithiothretol
EDTA	Ethylendiamine tetra acetic acid
EtBr	Ethidium Bromide
g	gram(s)
hr	Hour(s)
IPTG	Isopropylthio-β-D-galactoside
kb	Kilobase
LB	Luria Bertani medium
LTR	Long Terminal Repeat
M	Molar
mg	milligram (10^{-3}g)
min	minute
ml	milliliter (10^{-3}l)
mM	millimolar (10^{-3}M)
ng	Nanogram (10^{-9}g)

nMoles	nanomoles (10^{-9}M)
OD	optical density
ORF	open reading frame
PAGE	Polyacrylamide gel electrophoresis
PCR	Polymerase chain reaction
pMoles	picomoles
QTL	Quantitative trait Loci
rpm	Revolution per minute
RT	Room temperature
SDS	Sodium dodecyl sulphate
Sec	Second(s)
T	Thymine
TAE	Tris acetate EDTA
TBE	Tris borate EDTA
TEMED	N, N, N, N^1,- tetra methylethylenediamine
U	Unit
UV	ultraviolet
v/v	volume/volume
w/v	weight/volume
X-gal	5-Bromo-4-Chloro-3-indolyl-β-D-galactoside
µCi	Micro curie (10^{-6} Ci)
µg	microgram (10^{-6} g)
µl	Microliter (10^{-6} l = 10^{-3} ml)
µM	Micromolar (10^{-6} M)

Contents

Contents

Chapter 1: Introduction

Stability has been considered one of the hallmarks of all genetic materials. This stability could be broken by genetic recombination and mutations. However, following advent of molecular techniques, it has come to the realization that the genomes are not as stable as were thought earlier but contain a certain degree of fluidity. This fluidity in large part is contributed by the presence of genetic elements that are capable of moving from one location to another within the genome. Such elements are called transposable elements or mobile genetic elements. Barbara McClintock discovered them more than 50 years ago in maize, and she named them as controlling elements. McClintock proposed their major role in evolution because transposable elements are a source of hypermutagenicity. Now they are considered integral constituents of all the genomes from prokaryotes to eukaryotes. In eukaryotes transposable elements make up to > 40% of the total nuclear DNA (Human Genome Project). This percentage may exceed up to 60% in more complex genomes. Transposable elements have played a very important role in genome evolution by rewriting it, whenever needed.

On the basis of their structure and mechanism of transposition, mobile genetic elements are classified into two categories: Class I and Class II elements. The Class II elements, called transposons transpose by "cut and paste" mechanism catalyzed by element-encoded enzyme, transposase. Structurally they contain inverted terminal repeats (TIRs) and cause target site duplication upon integration. Depending on their functionality, they may be autonomous or non-autonomous elements. An autonomous element can transpose by itself or can mediate the transposition of its cognate non-autonomous elements. Non-autonomous elements are deletion derivatives of autonomous elements and may be defective in transposase activity. A number of class II elements have been characterized from a diverse group of plant species such as maize, *Antirrhinum*, petunia, *Arabidopsis* etc. The elements have also been shown to transpose in heterologous plant systems. They have extensively been

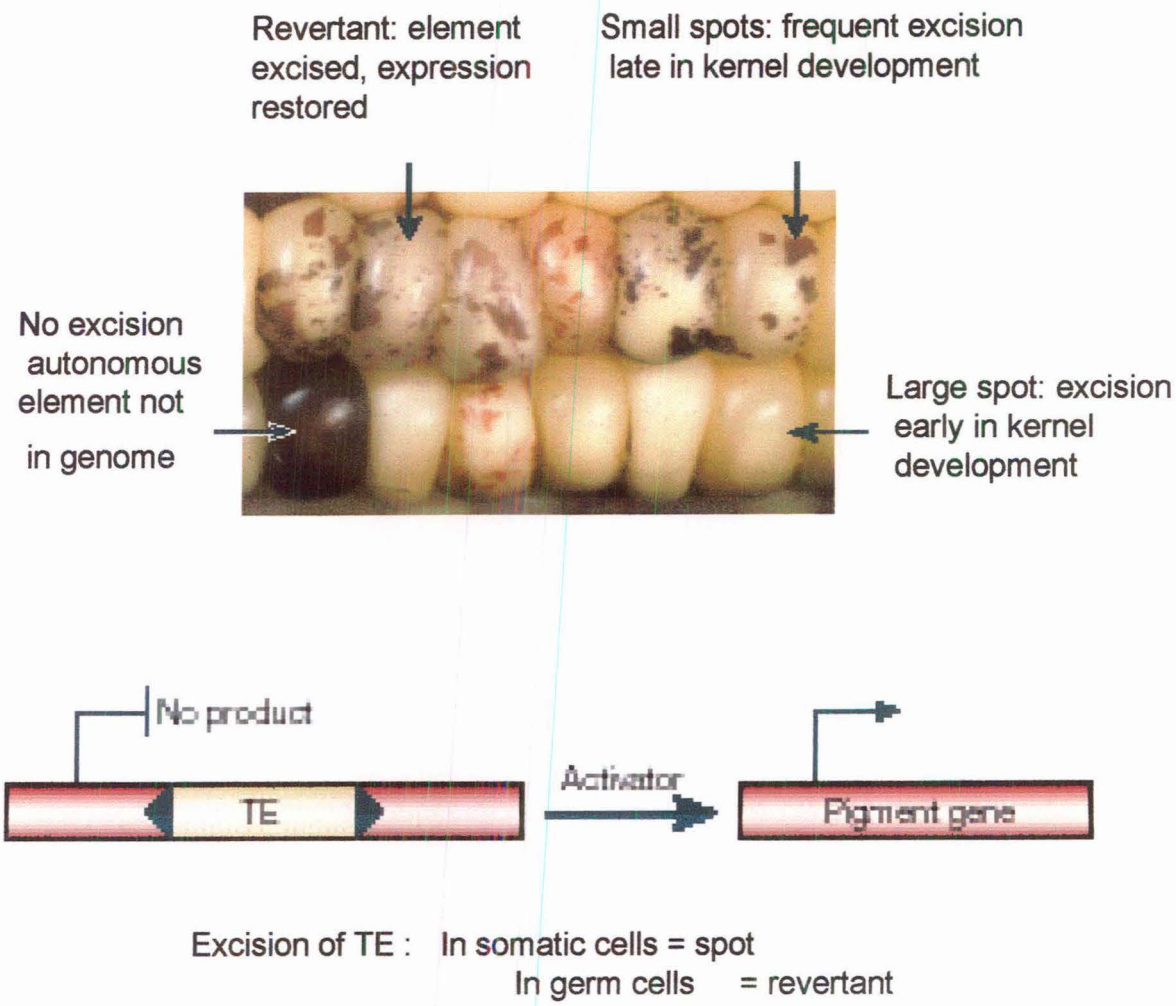


Figure:1 Using kernel phenotypes to study transposon behaviour. Kernels on a maize ear show unstable phenotypes due to the interplay between a transposable element (TE) and a gene that encodes an enzyme in the anthocyanin (pigment) biosynthetic pathway. Sectors of revertant (pigment) aleurone tissue result from the excision of the TE in a single cell.The size of the sector reflects the time in the kernel development at which excision occurred. An understanding of the genetic basis of this and similar

mutant phenotypes led to the discovery of TEs (Feschotte and Wessler, 2001)

Table 1 **Examples of transposable elements in plants**

Class/Subclass/ number of Superfamily family	Species	Autonomous mamber(s)	Non-autonomous member(s)	Copy the entire
Class 1				
Non-LTR Retrotransposons				
LINEs; L1 clade	*Z. mays*	Cin4	-	50-100
	L. speciosum	Del2	-	250,000
	A. thaliana	Tal1	-	1-6
SINEs	*N. tabacum*	-	Ts	50,000
	B. napus	-	S1	500
LTR Retrotransposons				
copia-like	*N.tabacum*	Tnt1	-	>100
	N.tabacum	Tto1	-	30-300
	Hordeum sp.	BARE-1	-	5,000-22,000
	O.sativa	Tos 17	-	2-30
	Z.mays	Hopscotch	-	5-8
	Z.mays	Opie-2	-	100,000
	Z.mays	-	BS1	1-5
Gypsy-like	*Z.mays*	Magellan	-	4-8
	Z.mays	Huck-2	-	200,000
	O.sativa	RIRE2	Dasheng	1,200
	A.thaliana	Athila 4	-	22
	A.thaliana	Athila 6	-	11
	A.thaliana	Ta3	-	1
	A.thaliana	Tar17	-	2
Class 2				
DNA transposons				
hAT	*Z.mays*	Ac	Ds	50-100
CACTA	*Z.mays*	spm	dSpm	50-100
	A.thaliana	CAC1	CAC2	4-20
Mutator	*Z.mays*	MuDR	Mu1	10-100
	A.thaliana	AtMu1	-	1-4
PIF/Harbinger	*Z.mays*	PIFa	mPIF	6,000

used for isolation of genes by transposon tagging. Besides these transposons, miniature inverted repeat transposable elements (MITEs) have also been found to exist in large copy numbers in plant genomes. MITEs are small elements ranging from 100 to 500 bp, and show preference for insertion into 2 to 3 bp A and T rich sequences targets.

Class I elements or retrotransposons transpose via RNA intermediate by "copy and paste" mechanism. Structure of these elements resembles sequences of retroviral genomes. Depending on the presence or absence of long terminal repeats (LTR) they are grouped into two categories: LTR or non-LTR-retrotransposons. Internally, they contain "gag" (group antigen), endonuclease and reverse transcriptase domains. Depending on the internal arrangement of RT and endonuclease domains, the retroelements have been named *Ty*1-*copia* or *Ty*3-*gypsy* family of retrotransposons. In *copia*-like retrotransposons the endonuclease domain is positioned 5′ to the reverse transcriptase domain, while in the gypsy type; it is at the 3′ end of the reverse transcriptase domain, but functionally both are similar. The two LTRS are very similar in sequences and the 5′ LTR contains promoter used for the transcription of the element. Besides these two other classes of retroelements called LINEs (Long Interspersed Nuclear Elements) and SINEs (Short Interspersed Nuclear Elements) have been found to exist.

Recently, a unique and previously unknown group of DNA transposable elements, named *Helitrons* has been identified in the genome of model plant *Arabidopsis* and worm, C*aenorhabditis elegans* (Kapitonov and Jurka, 2001; Feschotte and Wessler, 2001).This autonomous element transposes via rolling circle replication and also encode 5′→3′ DNA helicase; nuclease or ligase similar to encoded by rolling circle replicons. TIRs are absent but it transposes efficiently between 5′ – A and T-3′ with no extra modification of the AT sequence rich target sites.

Retroelements are widely distributed in plant kingdom. A large proportion of plant genomes are now known to consist of retroelements.

For example maize genome contains over 60% of its sequences as retroelements. The "c-value-paradox" i.e. non-correspondence between structural complexity to functional complexity could largely be explained by the proportion of retroelements. The larger genomes contain large proportion of retroelements. These elements have also been implicated in genomic expansion during evolution.

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. Besides they are also used in transposon tagging. The retroelements are dispersed through the genome and the sequences are relatively stable. Retrotransposons are now being utilized as molecular markers in DNA finger printing, genetic linkage mapping and phylogenetic analyses.

Chickpea (*Cicer arietinum* L) is the most important legume crop in the Indian subcontinent and ranks third in the world for pulse production today. This self-pollinating annual diploid crop with a somatic chromosome number of 2n = 16 has a genome size of 738 Mb. Such a large genome is expected to have a significant fraction of retroelements. Legumes such as chickpea have a narrow genetic base and the techniques such as AFLP are not able to detect much polymorphism. Therefore, one has to rely on other techniques such as SSR or retroelement based markers. Due to their large proportion in the genome and their possible utility in DNA-based molecules, it is desirable to isolate and characterize retroelements from the chickpea genome.

The objectives of the work embodied in this dissertation are

(i) Isolation of reverse transcriptase (RT) regions of the retrotransposons from chickpea genome using the RT-specific PCR primers.

(ii) Characterization of these sequences with respect to their complexity and sequence divergence.

The sequences would subsequently be used for isolation of retrotransposons from the chickpea genome.

Chapter 2: Review of Literature

2.1 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 total DNA) 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 retroelements discovery in plants started after Drosophila (e.g. *copia* and *gypsy*) and yeast (e.g. *Ty*1 and *Ty*3) (Shepherd *et al.,* 1984). Retrotransposons are indeed the most abundant and widespread class of transposable elements in plants (Table:1). 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 their 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 *Ty*1-*copia* group, because of frequent use of Ty1-copia specific primers to amplify and clone *Ty*1-*copia* relatives (Flavell *et al.,* 1992). However, *Ty*3-*gypsy* group of retrotransposons are also well represented. Some examples of the *Ty*1-*copia* group of retrotransposons are *BARE*1 (barley), *Bs*1, *Opie*, *PREM*-1 (maize), *SIRE*-1 (soybean), *Tnt*1, *Tto*1 (tobacco), *Tos*17 (rice), *panzee* (chickpea); and those of the *Ty*3-*gypsy* family are *Athila* (*Arabidopsis*), *cereba* (barely), *cinful*, *Grande*-1, *Zeon* (maize), *RIRE*3 (rice), *cyclops*-2 (pea). The non-LTR retro-elements LINE and SINE

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). Initially these were observed during sequence analysis of a genomic clone containing the potato urease gene. TRIMs are said to involved in restructuring plant genomes (Kumar *et al.*, 2001).

2.2 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. 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 (fig: 2). 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 (kilobases) 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. They have been used as DNA markers to study biodiversity in

Figure:2 Strategies for different types of retrotransposon based molecular systemsare shown both LTR and non-LTR retrotransposons.Retrotransposons are classified into two types, those with long terminal repeats(LTR) and those without LTR . LTR retrotransposons further sub-classified into Ty3-gypsy groups. Non-LTR retrotransposons consist of long interspersed repetitive elements(LINE) and short interspersed repetitive elements(SINEs).

- Sequence specific amplification polymorphism (S-SAP)

(b) retrotransposon internal variation polymorphisms(RIVP)
(c) inverse retrotransposon amplified polymorphisms(IRAP)
(d) retrotransposon-microsatellite amplified polymorphism(REMAP)
(e) retrotransposon-based insertion polymorphism(RBIP)

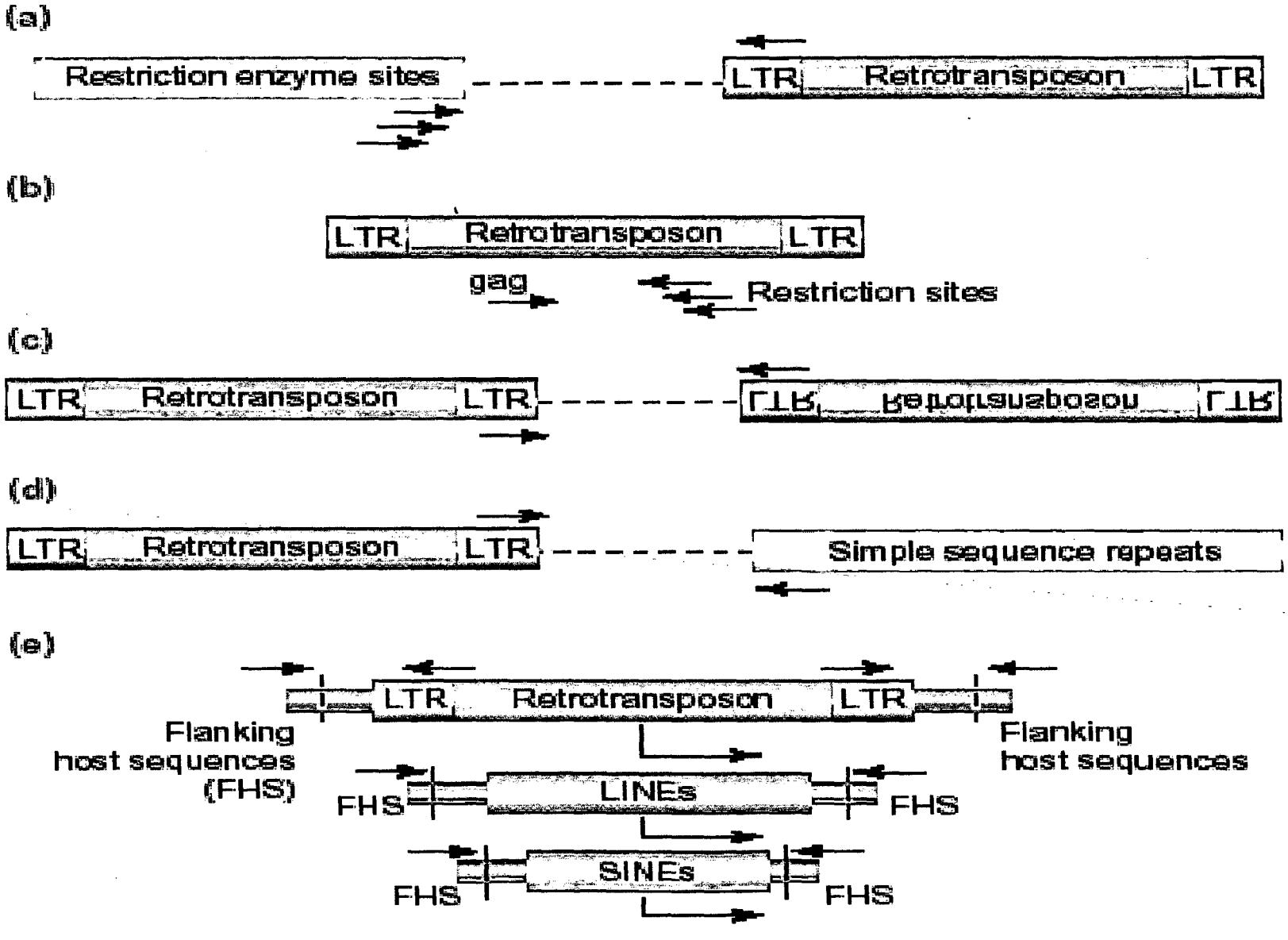


Figure:2 Strategies for different types of retrotransposon based molecular marker systems (Kalendar *et al.*,1999)

maize, pea and barley (Purugganan and Wessler, 1995; Ellis *et al.*, 1998; Kalendar *et al.*, 1999) 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:

(i) **S-SAP (Sequence-Specific Amplified 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 designed from the conserved terminus of the LTR and one based on the presence of a nearby restriction endonuclease site;

(ii) **IRAP (Inter-Retrotransposon Amplified 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. Fragments are separated by high-resolution agarose gel-electrophoresis (Kalender *et al.*, 1999).

(iii) **REMAP (Retrotransposon-Microsatellite Amplified 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);

(iv) **RBIP (Retrotransposon Based Insertional Polymorphism)**: This is codominant marker system that uses PCR primers designed from the retrotransposon and its 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).

SSAP markers, based on retrotransposons, have been used to construct linkage maps in barley (Manninen *et al.*, 2000), oat (Yu and Wise, 2000) and pea (Ellis *et al.*, 1998) using the LTR specific sequences (Pearce *et al.*, 1999). 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 much 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. Other examples are the M^a locus in barley, the *Rps1k* in soybean, *syn*19 and *piz* resistance genes to *Magnarporthe grisea* in pea, and the *Hero* locus conferring resistance to potato cyst nematode in tomato (see Kumar and Hirochika, 2001, for a review).

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 then be detected and used to establish the temporal sequences of insertion events, helping to determine phylogenies. These genetic properties have recently 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 recently to

estimate phylogenetic relationships between species in legumes (Pearce *et al.*, 2000) and cereals (Gribbon *et al.*, 1999; Kalender *et al.*, 2000).

2.3 Family and Structure of Retroelements

There is very much variations in members of retroelement family that contain reverse transcriptase encoding gene, which catalyzes the reverse transcription of DNA molecule from the RNA template. This reverse transcriptase or RNA dependent DNA polymerase was discovered around 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); (ii) without LTRs but have retroviral like *gag* and *pol* genes. 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. The relationship between various retroelements has been established and a phylogenetic tree has been constructed, based on the seven peptide regions (domain 1-7) spanning 178 amino acids which are common to all retroelements.

Phylogenetically when viral RNA polymerase sequences are used to root the tree, all rt containing elements falls into two major branches. One branch comprised of the bacterial msDNAs, group II introns and non-LTR retrotransposons while the other branch contains the three types of viruses

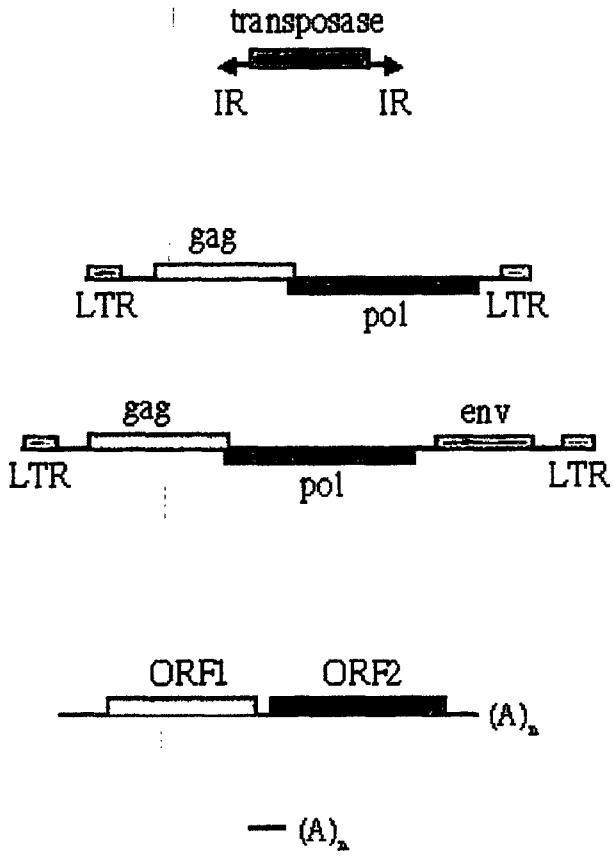
DNA transposons

LTR retrotransposons

Retroviruses

Non-LTR retrotransposons (LINEs)

SINEs (non autonomous)



Eukaryotic transposable elements

IR : inverted repeats ; LTR : long terminal repeats.

Figure:3
Different types of transposable elements.

(HBVs, CaMVs and retroviruses) and the LTR-containing retrotransposon groups (*Copia* and *Gypsy*).

LTR retrotransposons contain, long terminal direct repeats of generally 200-500 nucleotides that flank a region of 4.5 kb to 9.0 kb. 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 polyproteins that are later cleaved into low-molecular weight proteins for functionality reasons. The polyproteins encoded by *pol* is processed into (i) reverse transcriptase, which reverse transcribes RNA into double stranded DNA; (ii) aspartic protease, which cleaves the poly proteins into their component proteins (sometimes encoded by the *gag* gene); (iii) RNaseH, responsible for RNA template degradation for insertion of double stranded DNA copy of the element into the host genome. The third ORF, *env* encodes a membrane spanning protein that is intimately involved with infectivity in retroviruses. Although homologous *env* ORFs, are apparently not present in retrotransposons, but sometimes *env*-like ORF being present may encode analogous membrane spanning proteins, leading to doubt about infectious nature of retrotransposons (Granalbastein, 1992).

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 suggest that some of these are acquired sequences from other sources (Palmgren, 1994). Also, there exist 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 (Feschotte et al., 2002 review).

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 et al., 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. Work on vertebrates (SINEs account upto 5% of the genome) have 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′ end is quite variable in length and 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 As. 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 (Bennetzen, 2000 review).

2.4 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: 4). 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: 5). Transcription of LTR rerotransposons 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 (Bennetzen *et al.,* 1996; Grandbastien, 1992). Although no specific integration site is reported but the insertion is non-random. 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). 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 *Ty*1 and *Ty*3 elements insert primarily near tRNA genes, while TART of *Drosophila* specifically integrates at Chromosome termini. Gel blot and *in situ* hybridization analyses have indicated that many retrotransposons are dispersed throughout their host genome except in cenetromeric regions (Bennetzen, 1996). However, Pelissier *et al,* (1995), have shown that despite the presence of numerous euchromatic copies, the *Athila* element of

Figure:4 Mechanism of Retrotransposition and Retrotransposon Integration

Reverse transcriptases from retroelements. **a** | 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. **b** | Sequence alignment of the primary structure of the catalytic core of the reverse transcriptases from various retroelements. There are seven conserved regions (grey) situated in the 'finger' and 'palm' domains33. The positions of the catalytic residues involved in DNA polymerization are indicated by stars. All reverse transcriptases from LTR retroelements and retroviruses have a compact organization (top part of figure), with short 'spacer' sequences among the conserved domains (in yellow). Conversely, reverse transcriptases from non-LTR retroelements have the organization shown in the bottom part of the figure. The bar represents 50 amino acids (Kumar et al., 2001)

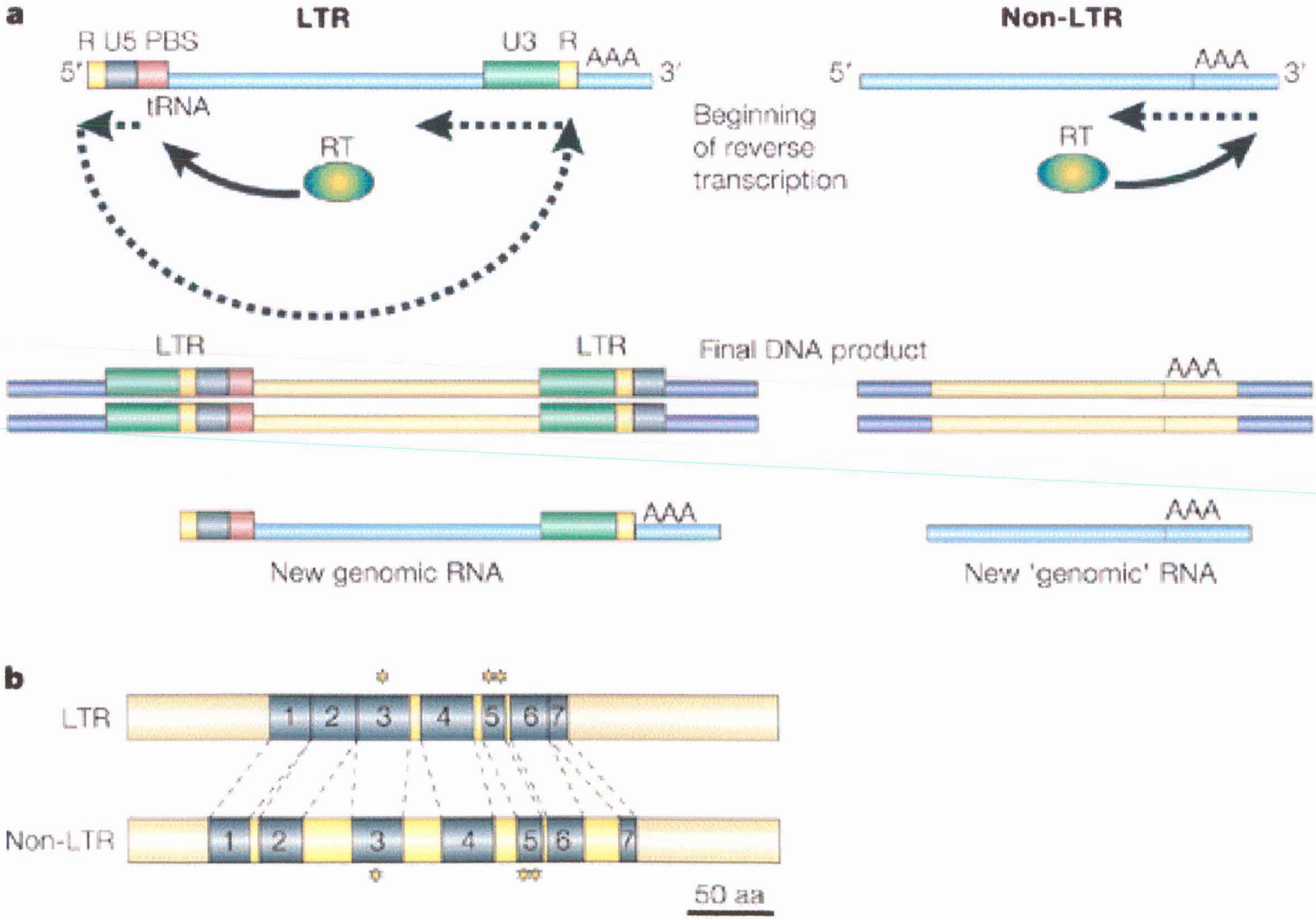


Fig:4 Mechanism of Retrotransposition and Retrotransposon Integration (

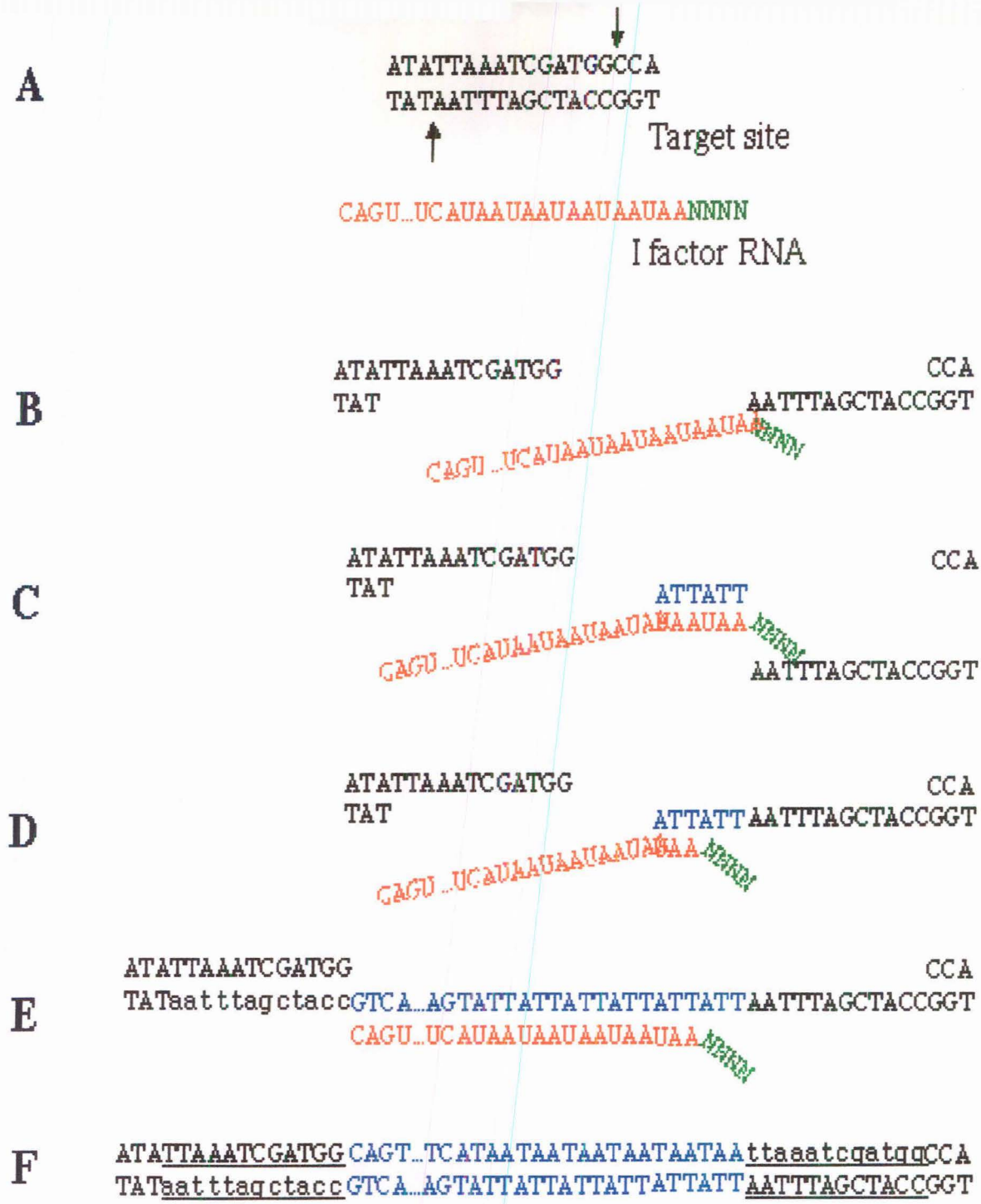


Fig: 5 Integration of retrotransposon into genome.
(A) The endonuclease cleaves chromosomal DNA at the target site for integration (arrows). (B) The RNA intermediate associates with the target DNA. (C) Reverse transcription initiates using the chromosomal DNA as primer and RNA as template, starting with in the UAA repeats. (D) The RNA template dissociates from the chromosomal DNA and anneals to the newly synthesized sequence. (E) Reverse transcription extends to complete first DNA strand synthesis. (F) After degradation of RNA by RNase H, second strand synthesis is completed. This produces a new retrotransposon. The target DNA is in black, the RNA intermediate in red with extra nucleotide flanking in green, and the cDNA in blue. The target site duplication is underlined (Wilhelm, 2001)

Arabidopsis is concentrated in or near heterochromatic regions and have provided strong evidence that most of the heterochromatic elements retrotransposed directly into 180bp satellite clusters.

Several elements (*PREM*1, *Grande*1 and *Cinful*1) 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 suggests that Cin4, *Ta*1 and *Tnt*1 are like Ac and *Mu*1 in their preferential association with genes containing isochors (Bennetzen, 1996).

Plant retroelements are also found in mitochondrial DNA of *Arabidopsis* especially *Ty*1-*copia* and *Ty*3 *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 analysis shows that 5% of the Arabidopsis mitochondrial genome appears to have retro-elemetnal origin.

Like those of other eukaryotes, fungal retrotransposons also transpose via RNA intermediate and employ reverse transcriptase for this purpose (Kempken and Kuck, 1998). Moran, J.V reported that LINE-1 (L1) rerotransposition causes genomic deletions. Many times it resulted in the formation of Chimeric L1s, containing the 5′ and of an endogenous L1. Thus there is enough literature to 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 *Ty*1 retrotransposon are abundant, but new insertions are extremely rare, largely because only one *Ty*1 cDNA, on average, is made for every 14,000 *Ty*1 transcripts (Curcio and Garfinkel, 1999).

2.5 Diversity, antiquity, 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 *Drosophila* and yeast (Flavell *et al.,* 1992a). Even the extremely small *Ta* family of *copia*-like retrotransposons of *Arabidopsis thaliana* is highly diverse (Flavell *et al.,* 1992b).

Retrotransposons insertion in the 5′ and 3′ regions, near normal genes, coupled with their degenerate nature provides additional evidence for an ancient association between *copia*-like retrotransposons and plant genomes. Maize gene *adh*1 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). 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 *Ta*1 in *Arabidopsis* and of *Tst*1 in *Solanum tuberosum* to >40,000 of *del*1 in *Lilum longiflorum* is observed (Grandbastien, 1992). Many other retrotransposons have been identified as extremely abundant sequences in plants. 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 *BARE*1, *BIS1* and *del*1 are present per haploid genome and they account for significant fraction of their host nuclear genomes (Bennetzen, 1996). The 4.45 kb LINE *del*2 has 250,000 copies in *Lilium speciosum* comprising 4% of its genome (Wessler *et al.,* 1995).

2.6 Retroelements and C-value Paradox

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 may be defined as the absence of correlation between increases in DNA content and complexity of an organism i.e. genome size of an organism does not depends 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 LTR retrotransposons in the genome increases with its sizes, as is evident from rice, the smallest grass genome characterized (~15%. of its 430 Mb genome consist of LTR retrotransposons (Wessler *et al.,* 2001), maize with 50-80% retrotransposons of ~2,800 Mb genome (Bennetzen and SanMiguel, 1998) (Morgante *et al.,* 2001) and barley with > 70% retrotransposons of ~ 4800 Mb genome (Vicient *et al.,* 1999). 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 have nearly 14% transposon and retrotransposon sequences (*Arabidopsis* Genome Initiative, 2000). Similarly, study of a 211kb contiguous genomic region of *Triticum monococcum* orthologus to the Lr10 leaf rust resistance locus in hexaploid wheat reveals an overall gene diversity of one

gene per 42kb and 70% of these sequences comprise several classes of transposons and retro-elements (Wicker *et al.,* 2001).

2.7 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 deletion, rearrangements and/or replacements (Messing *et al.,* 1995). This predominance of defective elements is partly due to the self-mutagenic properties of the DNA elements (Kunze *et al.,* 1996), but is also likely to be associated with an intrinsic higher mutation rate of cytosine-methylated DNA. 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 *et al.,* 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. 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 and Fagard, 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 ddm*l (decrease DNA methylation1) 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 *ddm*l mutation have notably decreased CpG methylation. Consequently, the transcriptional derepression of TEs in the mutant strains was accompanied by demethylation of the elements. The *ddm1* gene encodes a protein with strong similarity to SWI2/SNF2 Chromatin-remodeling factors (Richards *et al.,* 1999). These results indicate a possible function link between chromatin remodeling, DNA methylation and genome integrity.

Reactivation of TEs *ddm*l 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.

2.8 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 of affairs for most of the thousands of plant retrotransposons present in a given plant is virtually undetectable rates of transposition. They are rarely (if ever) active during normal plant development. In contrast, retrotranscripts can be quite abundant in yeast (Curcio *et al.,* 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) often producing transcripts that can only be detected by amplification techniques (Hu *et al.,* 1995). Transcripts homologous to some elements are found at relatively high levels and/or in some tissues. However, element-dependent expression has only been convincingly demonstrated in a few cases (Bennetzen, 1996). Even in the cases where element specific transcriptional initiation is observed, the low level of RNA detected suggests that only a small subset of the element population is being expressed. During study of expression of components of the yeast retrotransposon *Ty*1 in *E.Coli*, it was found that polypeptides encompassing the capsid-forming component of *Ty*1 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 article

Figure:6

Detection of new genomic insertions by transposon display. The transposon display technique for detecting new transposable element(TE) insertions is illustrated through the use of an example of stress-induced mobilization of the Tnt1 retrotransposon in tobacco.

a)The isolation of protoplasts from tobacco leaf cells activates a stress response that induces transcription of the Tnt1 LTR retrotransposon.Induction might be mediated by the binding of a transcriptional activator to a *cis*-motif(yellow box) in the LTR. Some Tnt1-encoded mRNAs are converted into double stranded cDNAs that integrate into the tobacco genome.

b)The transposondisplay protocol begins with the digestion of genomic DNA with a restriction enzyme(here EcoRI). Most of the thousands of restriction fragments do not have insertions. Adaptors(pink boxes) are ligated to all fragments and the mixture is used as template for one(as shown) or two(not shown) round(s) of PCR amplification, using primers that are complementary to subterminal TE sequences(arrows)

New integration events appear as additional bands(red arrows) on an autoradiograph of a transposon-display gel of genomic DNA. In this example, DNA was isolated from tobacco plants either before treatment(lane 1,6) or after regeneration from protoplasts(lanes 2-5,7). (Feschotte and Wessler, 2001)

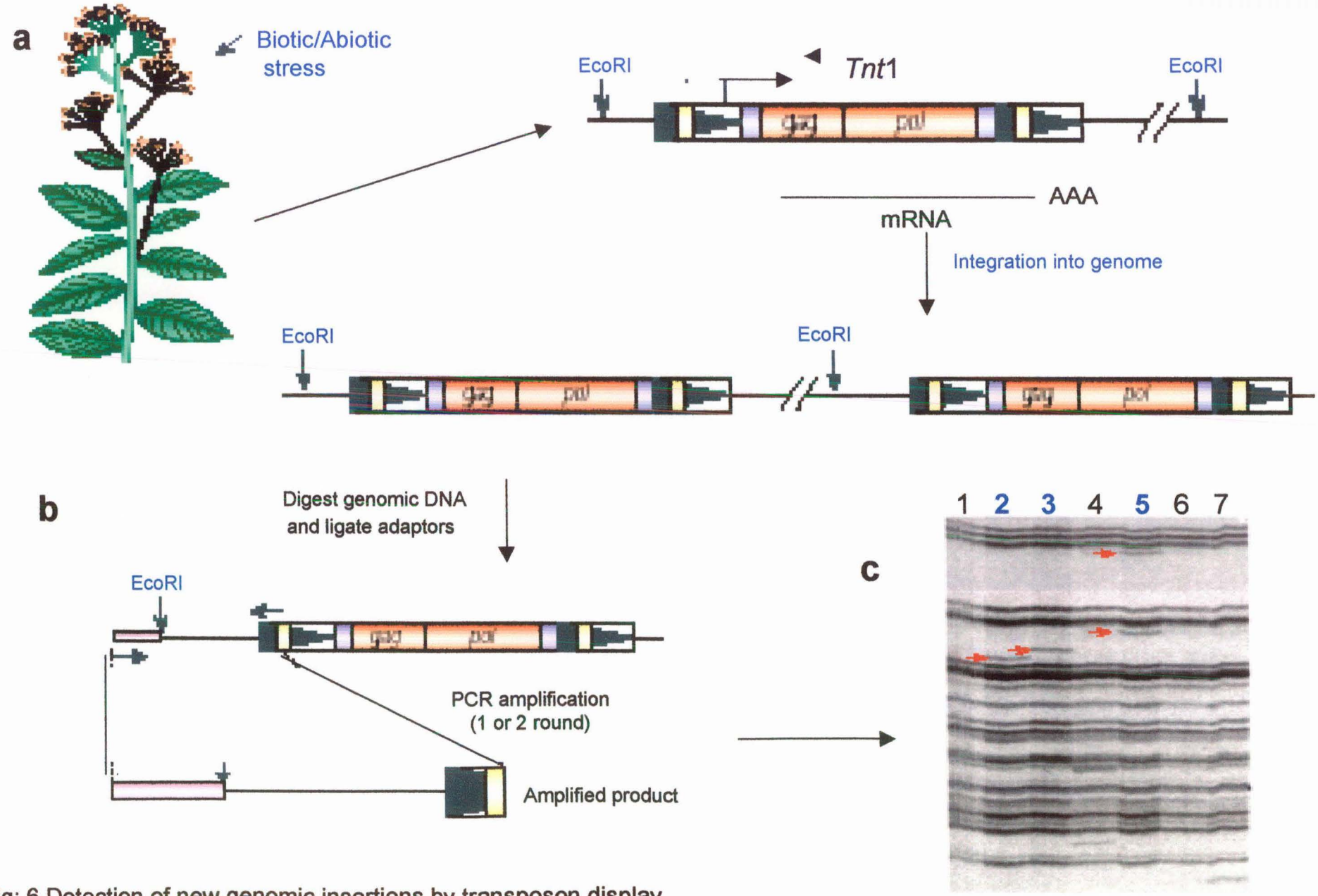


Fig: 6 Detection of new genomic insertions by transposon display

supports that these steps are independent of host factors (Luschnig and Bachmair, 1997). The *gag* homologue of *Ty*1 assembles into spherical particles similar, but not identical to virus like particles in the natural host of *Ty*1, *saccharomyces cerevisiae*. Assembly process fails upon deletion of a domain in the C-terminus of reading frame of *Ty*1. Also, *Ty*1 *gag* fusion proteins can be produced in *E.Coli* an organism lacking endogenous retrotransposon (Bachmair, *et al.,* 1995). 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). Many of the plant retrotransposons studied so far are transcriptionly activated by various biotic and abiotic stresses (wessler, 1996; Mhiri *et al .,* 1997; Beguiristain *et al.,* 2001; Fig 6). Expression of the tobacco *Tnt*1 and *Tto*1 retrotransposons is greatly enhanced by several abiotic stresses including protoplast isolation, cell culture, wounding, methyl jasmonate , $CuCl_2$, UV and salicylic acid (Takeda *et al.,* 1998, 1999). The same is true with *BARE*-1 and *OARE*-1 retrotransposon of barley and oat respectively (Vicient *et al.,* 1999; Schulman and Nevo, 2000; Nakayashiki and Mayama, 2001). Similarly, biotic stresses, such as infection by viral, bacterial and fungal pathogens, can activate the transcription of these elements (Grandbastein *et al.,* 2001), as shown of *Tnt*1 by fungal factors. *Tnt*1 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 *Tnt*1 elements is also induced when tobacco cells are treated with fungal extracts containing cell wall hydrolases (Pouteau *et al.,* 1991).

In contrast to *OARE*1 and *Tnt*-1, transcription of *Tos*10, *Tos*17, *Tos*19 in rice and *Tto*1, *Tto*3, in tobacco is induced by cell culture and protoplast culture respectively (Hirochika, 1993a, 1996). Plants regenerated from cultured tobacco or rice cells posses 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 *Tos*17 increased from 2 to more than 30 copies in some strains (Hirochika *et al.,* 1996; Grandbastin, 1998). Only a fraction of retrotransposon population in plants appear 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).

There is strong evidence that retroelements may confer certain selective advantage on the system possibly to withstand adverse environmental stresses. Analysis of *BARE*1 element of barley in "Evolutionary Canyon" Mount Carnel 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).

2.9 Origin of retro-elements

The presence of transposable elements in all living organisms suggests 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 rerotransposons 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). 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 *BIS*1 elements is lacking or reduced in centromeres telomeres and nucleolar organizer (Moore *et al.,* 1991), while the Grande1 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.

Recently originated a very 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). These types of transfers might have occurred into injured tissues. This model favours origin of retrotransposons in 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). Thus, some defective retroviruses might have given rise to LTR-retrotransposons (Bennetzen, 2000) because recently defective appearance of the *env*-derived regions in LTR-retrotransposons have been reported in many plants (Gaucher *et al.,* 1998).

2.10 Retrotransposon Evolution and Relation with 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).

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, retrotransposon use the same proteins for their replication as are used by some integrating retroviruses. So it is thought that some retrotransposons are similar to lysogenic retroviruses. 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*) envelop protein coding gene necessary for intercellular transmission. 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 then *copia* and retroviruses. Thus *gypsy* could be considered for infectious nature of some retrotransposons (Sinkovics, 2001).

Retroviruses can acquire and transmit portion of their host genome by process of transduction. Viruses can take up protoncogenes 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 suggest 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 *Bs*1 should be defective version of still undetected plant retroviruses. The sequence of *Bs*1 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 *Bs*1 has acquired transmembrane domains of a maize proton ATPase gene, *Mha*1 and has selected for the conservation of the reading frame and 'amino-acid sequences of this acquired segment (Jin and Bennetzen, 1989). Also *Bs*1 belongs to *Ty*3-*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 *Bs*1 has primarily undergone conservative mutations since acquisition from *Mha*1, suggesting a selection for function. The portion of *Mha*1 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. Research are in progress to find the origin of these internal sequences, because 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 retro-elements 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 (Sinkovis, 2001).

2.11 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 dehydrogenase*1 gene (*adh*1), is present in only 1-5 copies in the maize genome (Johns *et al.,* 1985).

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 al.,* 1998) and serving as molecular markers 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. Insertion of retrotransposons have shown either inactivation or alteration in gene function (wessler *et al.,* 1995). Nearly 80% mutations

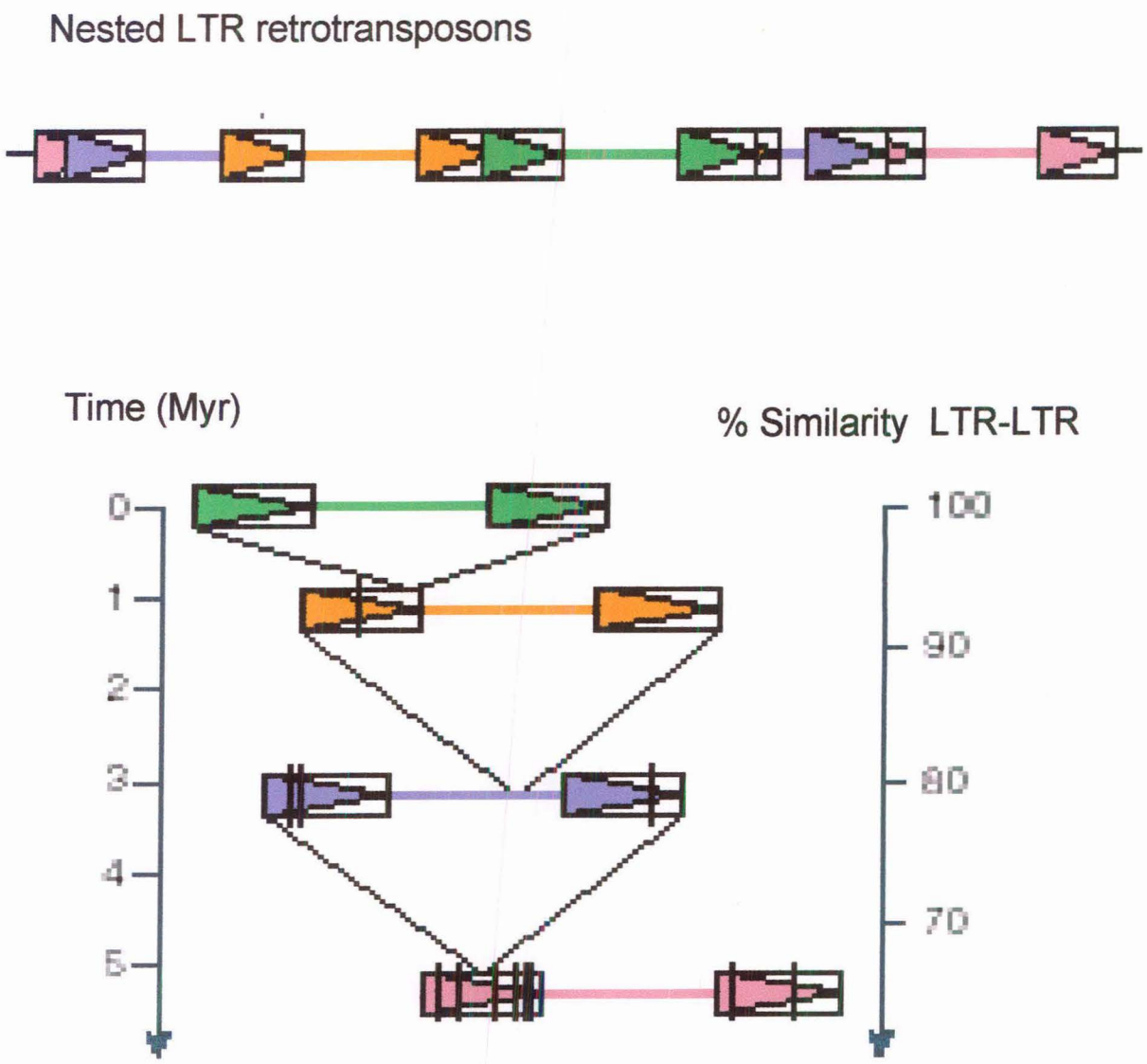


Figure:7 Estimating the time of retrotransposon insertion. At the time of insertion, The long terminal repeats (LTRs) of an element are identical because both are copied from the same template during cDNA synthesis. As time pases, nucleotide changes accumulate in each LTR (represented by vertical bars in the LTRs). If the average rate of nucleotide substitution per year is known for the host organism, then sequence divergence between the LTRs provides an estimate of when reaction occurred. This method has been applied to date the insertions of LTR retrotransposons nested in the intergenic regions that surround the maize alcohol dehydrogenase gene. Myr-million years. (Feschotte and Wessler, 2001)

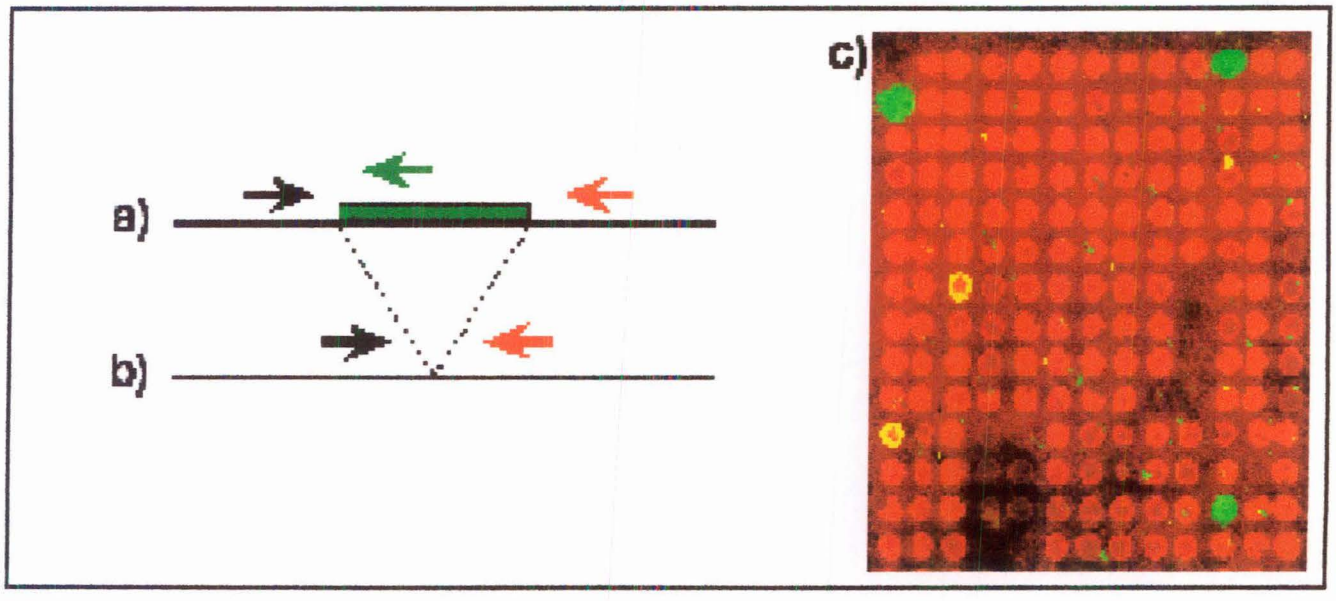


Figure:8 Detection of presence or absence of retrotransposon insertions by PCR.
The polymerase chain reaction (PCR) is used to detect the presence
• or absence (b) of a transposon insertion. The two different PCR products can be labelled with different fluorescent molecules, which can be detected by microarray technology (c). In this case a red fluorescent spot indicates transposon absence, a green spot indicates presence, a yellow spot shows the presence of both alleles and a gap shows the absence of both alleles (Flavell *et al.*, 2000)

detected in *Drosophila* caused due to transposition. 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 event 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 enkaryotes, most recombination is limited to genes, thus limited ectopic recombination is reported (Bennetzen et al., 2000 review).

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 1000000 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). 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 (Thomas, 1971). 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 (~15% of its 430 Mb genome consist of LTR retrotransposons) (Feschotte et al., 2002), to maize (~ 2,800 Mb, 50-80% retrotransposons, San Miguel and Bennetzen., 1998; Meyers *et al.,* 2001) and barley (~ 4,800 Mb, > 70% retrotransposons, Vicient *et al.,* 1999).

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 (Fig: 7 and 8). Their initial observations was dramatically followed up with the demonstration that bursts of LTR retrotransposon activity have doubled the maize genome with in the past 6 million years (Myr; San Minguel *et al.,* 1998; Bennetzen *et al.,* 1996). The temporal components to their analysis was 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. 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 (Kumaret 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 (Grandbastein et al., 1998; Kalendar et al., 2000).

Chapter 3: Materials and Methods

3.1 Materials

	Source
Seeds of *Cicer arietinum*	Variety Pusa 362 (IARI)
Chemicals	
Acrylamide	Gibco-BRL
Agarose	FMC Corporation
Agar	Qualigens (Bacteriological grade)
APS	Sigma
Amberlite	Serva
Antibiotics	Sigma (Ampicillin)
Bis-Acrylamide	Gibco-BRL
Bromophenol blue	Sigma
CTAB	Sigma
Calcium chloride	Sigma
EDTA	Sigma
Ethanol	Merk
Ethidium bromide	Sigma
Ficoll	Sigma
GF/C filter	Whatman
dNTP	NEB
Phenol	Qualigens (Redistilled in the lab)
Proteinase K	Sigma
Radioactive chemicals	^{35}SdATP BARC, India
Random-Primer DNA Labelling Kit	NEB
Restriction Enzymes	NEB
RNase A	Sigma
SDS	Sigma
Sequencing Kit	Sequenase version 2.0 kit USB/Amersham
Silane and γ-Silane	Sigma
Sodium hypochlorite	SRL
Taq DNA polymerase	Promega
TEMED	Sigma
Tris	USB
Tryptone	DIFCO
T4 DNA ligase	NEB
Urea	Gibco BRL
X-gal	Biosynth AG
X-ray films:	Kodak
Xylene Cyanol	Sigma
Yeast extract	DIFCO

All other common chemicals used were of the highest quality available and procured from local sources like Merk, Qualigens, BDH, CDH, SRL, Sd-fine chemicals.

Bacterial Strain	Genotype
E.coli strain DH 5α	recA1, endA1, gyr A96, thi-1, hsd 17 (r,m) supE44Δ, relaA1, deoR, lacU169 (Ø 80 lacZΔM15. NA resistance.
Vector	
pGEM T-easy	Promega
Plasmid mini prep-kit	Qiagen.

Methods

3.2 Preparation of Plant Material

The seeds of chickpea (*Cicer arietinum*) were grown in soil. As it is a rabi crop so seeds were grown in winter. Young plants (~10 cm in height) were taken except roots and 10 gm of tissue was snap frozen in liquid nitrogen after wrapping into aluminium foil. The frozen tissue was stored at -70^{0}C for later use.

3.3 Isolation of Genomic DNA from Plant Tissue

3.3.1. Extraction of Genomic DNA

For DNA isolation protocol used was modified version of Ausubel *et al* (1987). The plant tissue was carefully ground to fine powder using sterile mortar and pestle in the presence of liquid nitrogen. The fine powder of tissue was added to extraction buffer (100 mM Tris.Cl pH 8.0, 100 mM EDTA,250 mM NaCl and 100 µg/ml of proteinase K) contained in sterile GSA bottle. The bottle was gently swirrled to suspend the powder in extraction buffer. Sarkosyl added was 1% of total volume. The mix was incubated at 55^{0}C in water bath with gentle shaking for 1.5 hrs. Then the lysate was centrifuged at 6500 rpm at 4^{0}C for 10 min. In another sterile fresh GSA bottle supernatant was taken leaving debris. To the supernatant 0.6 volume of isopropanol was added and mixed gently by inverting the GSA tube to precipitate 15 min at 4^{0}C to get nucleic acid pellet. Pellet was washed with 70% ethanol two times, air dried and dissolved in TE.

3.3.2 Purification of DNA

To the DNA solution in TE, 5 M NaCl was added to adjust the final concentration 0.7 M and mixed thoroughly but gently. Ten percent CTAB solution containing 0.7 M NaCl was added to give a final concentration of 1% CTAB, mixed well and incubated at 65^{0}C for 10 min. It was extracted with

equal volume of chloroform: isoamyl alcohol (24:1) in a conical extraction tube and then was transferred to a sterile Oak Ridge tube and centrifuged at 12000 rpm for 10 min at 4^0C. With the help of wide bore pipette tip aqueous phase was taken and $1/10^{th}$ volume of 10% CTAB containing 0.7 M NaCl was added. Then chloroform: isoamyl alcohol step was repeated. The final clear aqueous phase was transferred to a sterile corex tube.

3.3.3 Precipitation of DNA

To the corex tube containing aqueous phase 0.6 volume of isopropanol was added to precipitate DNA. The tube was shaken gently till a white fibrous precipitate appeared. The precipitate was transferred to a tube containing 70% ethanol with the help of wide mouth pipette to avoid shearing, and pelleted by centrifugation at 8000 rpm for 15 min at 4^0C. The pellet was air dried and dissolved in TE.

3.3.4 Quantitation of DNA

Quantitation of DNA was done spectrophotometrically by measuring the absorbance at 260 nm of wavelength. The purity of DNA was checked by taking OD at 260 nm and 280 nm.

3.4 Isolation of Retroelements

3.4.1 Polymerase Chain Reaction

To amplify the reverse transcriptase region from the genomic DNA, polymerase chain reaction was set using degenerate oligonucleotide primers. The primers (26 mer and 27 mer) were diluted with sterile milliQ water to get the final conc. of 50 ng/µl each, which corresponded to 56.10 moles/µl for 27 mer and 58.27 pmoles/µl for 26-mer. The total volume of PCR reaction was set 25µl using 500 ng of genomic DNA, 1 mM of each dNTPs, 2.5 units of Taq DNA polymerase and 0.1 volume of 10x Promega PCR buffer. The concentration of $MgCl_2$ was adjusted 1.5 mM in one reaction and 2 mM in

other reaction. One to two drops of mineral oil was added to each tube to minimise evaporation. MJ Research thermal cycler was used for temperature cycling with following parameters: 94^0C for 5 min, following by 35 cycles of 94^0C for 1 min, 47^0C for 1 min and 72^0C for 1.5 min, followed by finally 72^0C for 10 min for extension. One fifth of the PCR product from each tube was electrophoresed on 1.2% agarose gel along with 100 bp DNA ladder.

3.4.2 Purification of Amplicons

The PCR amplified DNA was eluted from a 0.7% agarose gel in 1x TAE. The part of gel containing the desired band of DNA was cut with a sterile blade, chopped into fine pieces, placed in equal amount of Tris-equilibrated phenol and frozen at -80^0C overnight. The frozen mixture was taken out, centrifuged at 4000 rpm for 20 min at 4^0C and upper aqueous layer was extracted. To the aqueous layer one volume of chloroform: isoamyl alcohol (24:1) was added, mixed well and centrifuged at 4000 rpm for 10 min at 4^0C. Again aqueous layer was extracted, measured, DNA precipitated by adding 0.1 volume of 3M sodium acetate followed by 2.5 volumes of ethanol and stored at -20^0C for 30 min. This was followed by centrifugation at 4000 rpm for 15 min at 4^0C. The supernatant was discarded and the pellet was washed twice with 70% ethanol. The DNA pellet was air dried, dissolved in deionized milliQ water (10 µl) and concentration was determined spectrophotometrically.

3.5 Cloning of Amplicons

3.5.1 Vector

The commercially available pGEM-T Easy vector from Promega was selected. Since it has T at its both staggered ends, so PCR amplified DNA would easily be ligated due to presence of A at its both staggered ends, which are added during PCR reaction by Taq polymerase.

3.5.2 Ligation of Amplicons into the pGEM-T Easy vector

The ligation was done in Ix Promega ligation buffer, and volume of ligation-mix was adjusted to 10 µl. The molar ratio of vector to insert DNA was kept 1:3 by considering that the concentration of total DNA (insert + vector) must not exceed 30-40 ng/µl in the ligation mix. In the end 2 weiss unit of Promega T4 DNA ligase was added, mixed well and content was incubated at 4^0C for 20 hrs.

3.5.3 Preparation of Competent Cells

The cells of *E.Coli* DH5α strain were streaked on LB agar plate and incubated at 37^0C overnight to obtain single colonies. Single colony was picked up, inoculated in 25 ml of 2x medium (20% Bacto-tryptone, 10% Bacto-yeast extract, 1% NaCl and 4 ml 1 M NaOH per 100 ml, pH 7) and grown at 30^0C overnight with 175 rpm shaking. One ml of this overnight culture was inoculated in 100ml of 2x medium and grown at 30^0C with shaking (200 rpm) until OD at 600 nm wavelength was around 0.5. The flask was cooled on ice water mix for about 2 hrs and then the cells were pelleted at 3000 rpm for 5 min at 4^0C. After removal of supernatant, the cell pellet was gently resuspended in 25 ml of ice cold competent cell buffer (100 mM CaCl2, 70 mM $MnCl_2$ and 40 mM sodium acetate with pH 5.5), incubated on ice for 30 min and pelleted at 3000 rpm for 5 min at 4^0C. Supernatant was discarded and the cell pellet was gently resuspended in 2.5 ml of ice cold competent cell buffer. Autoclaved 1.15 ml glycerol (80%) was added drop wise with gentle shaking and the final content was aseptically aliquoted in 100 µl and 200 µl volumes in ice cold sterile vials. The tubes were stored at – 80^0C till further use.

To check the competency of the cells, 25 ng of supercoiled pUC19 plasmid (containing amp resistance gene) was added to 100 µl of competent cells and mix was incubated in ice for 30 min. One ml LB was mixed gently after a heat shock at 42^0C for 1.5 min. Then it was incubated at 37^0C for 1 hr at 200 rpm and plated on ampicillin (0.1 mg/ml) LB plates. The plates were incubated at 37^0C for 2 hr and colonies were counted.

3.5.4 Transformation

For each transformation a vial of 100 µl of competent cells was used. The frozen cells were allowed to thaw on ice. Three µl of ligation mix was added and mixed gently. The content was incubated on ice for 30 min, followed by heat shock at 42^{0}C for 90 sec and mixed with 1 ml LB gently. Finally incubated at 37^{0}C for 1 hr at 200 rpm and plated 100 µl on each LB plates (LB agar, 100 µg/ml ampicillin, 20 µg/ ml of x-gal and 100 µg/ml of IPTG). The plates were incubated at 37^{0}C for 12-16 hrs.

3.5.5 Screening of transformed Colonies

White colonies were selected, patched on LB plates and incubated at 37^{0}C for 12 hrs. The modified version of Birnboin and Doly (1979) was used for small scale plasmid isolation.

Each transformed colony was inoculated in 5 ml of LB (ampicillin 100 µg/ml) and incubated at 37^{0}C at 200 rpm overnight. One ml of culture was taken into sterile 1.5 ml Eppendorf tube and cells were harvested by centrifugation at 12000 rpm for 30 sec at room temperature and resuspended in 100 µl of solution I (50 mM glucose 10 mM EDTA and 25 mM TrisCl pH 8.0) by vortexing. The cells were lysed by adding 200 µl of freshly prepared solution II (0.2 N NaOH and 1% SDS). The contents were mixed by inverting the tube 5-6 times gently and was stored on ice for 5 min. To neutralize the Sol II, 150 µl of solution III (3 M potassium acetate pH 4.8) was added, mixed by inverting the tube rapidly 5-6 times and stored on ice for 5 min. The cell debris was pelleted by centrifugation at 12000rpm for 5 min at 4^{0}C, supernatant was transferred to fresh tube and was extracted with one volume of chloroform: equilibrated phenol (1:1). The mix was centrifuged at 12000 rpm for 2 min at 4^{0}C and to the supernatant two volumes of absolute ethanol was mixed, kept at room temperature for 2 min to precipitate DNA. The DNA was pelleted by

centrifugation at 12000 rpm, for 8 minutes at 4^0C, washed with 1 ml 70% ethanol, air-dried, dissolved in 50 µl of TE (pH 8) and stored at – 20^0C.

The plasmid DNA in TE was treated with RNaseA (Sigma,10 µg/50 µl plasmid DNA) at 65^0C for 15 min to digest RNA. Restriction digestion reaction was set up in 30 µl with 200 ng (at least) of plasmid DNA, 0.1 volume of 10x digestion buffer, 3 units of EcoRI restriction enzyme (NEB) was taken. The digestion mix was incubated at 37^0C overnight. The digestion was checked on 1.2% agarose gel along with 100 bp DNA ladder as a marker.

The plasmids of positive clones were isolated and purified by using Qiagen kits.

3.6 DNA Sequencing (Sanger's Dideoxy method)

The sequencing reactions were done by using sequenase version 2.0 DNA sequencing kit of USB.

Single stranded form of DNA is required for primer binding, so double stranded plasmid was converted into single stranded template form by alkali denaturation method. For denaturation, to 8 µg of plasmid DNA 1/10 volume of 2 M NaOH + 2 mM EDTA was added, incubated at 37^0C for 30 min and neutralized by adding 1/10 volume of 3 M NaoAc (pH 4.5-5.5). The denatured DNA was precipitated by 2.5 volumes of ethanol and pelleted by centrifugation at 13000 rpm for 20 min at 4^0C after keeping at -70^0C for 15-30 min. Pellet was washed in 70% ethanol, air deried, dissolved in sterile 14 µl milliQ and divided into two tubes equally (7µl each) for forward and backward reactions.

To the denatured DNA, 0.5 pmoles of primers (M13F and M13R) were added in Ix sequenase buffer in final volumes of 10 µl each for forward and backward reaction. After mixing the content, it was incubated at 65^0C and primers were allowed to anneal to the templates by cooling slowly to < 35^0C over time period of 30 min.

In labelling reaction, the annealed primers were extended using limiting concentrations of dNTPs and α^{35}SdATP. This step is thought to continue upto

complete incorporation of labelled nucleotides into DNA fragments which are distributed randomly in lengths. Five times dilution of labelling mix with sterile milliQ water and eight times dilution of DNA polymerase enzyme with ice cold enzyme dilution buffer was done from Sequenase version 2.0 kit for use in reaction. To the annealed template-primer, 1 µl of DTT (0.1 M), 2 µl diluted labelling mix, 0.5 µl α^{35}SdATP and 2 µl of diluted enzyme were added. The content was mixed uniformly and incubated at 20^0C for 2-5 min.

For termination of both forward and backward reactions steps are identical and can be done separately. Four Eppendorf tubes were labelled as ddA, ddC, ddG, ddT and termination mixes (2.5 µl each) ddATP, ddCTP, ddGTP, ddTTP were taken in them respectively. These tubes were pre-warmed to 37^0C and 3.5 µl of labelled mix was added to each termination mix tube, which was followed by incubation period of 5 min at 37^0C. Finally 4 µl of stop solution was added to each tube, mixed well and stored at -20^0C.

Sequencing plates were carefully cleaned using detergent and finally rinsed with distilled water. Both plates were siliconised. Siliconization of plates was done by wiping with chloroform, followed by 5% dimethyl-dichlorosilane in chloroform and finally with ethanol. The plates were assembled with their siliconised surfaces facing each other with the spacers held between them on lateral side with the help of clamps. Leakproof tape was applied to bottom side to seal the gap. Six percent acrylamide-area gel was poured into the gap between plate to form gel.

Acrylamide-Urea gel (6%) was prepared as follows: 5.7 g of acrylamide, 0.3 g of bisacrylamide, 0.6 g of amberlite were added in 15 ml of distilled water, stirred for 2 hrs. and filtered through Whatman no.1 filter, to the filterate added 45 g urea, 10 ml of sterile TBE and volume was raised to 100 ml with distilled water. The solution was filtered through whatman no.1 filter, added 0.1 g ammonium persulphate and 9 µl TEMED. Immediately it was poured into assembled plates held at an angle of 45 degrees using syringe with needle. This was done in a fashion to avoid trapping of any air bubble in the

gel. Then the plates were kept in horizontal position and flat surface of the shark tooth comb was inserted in gap between plates at the top up to a depth of 4 mm inside the acrylamide gel solution. The plates were left in that position overnight without disturbance for acrylamide polymerization.

3.7 Electrophoresis of Sequencing gel

The BRL sequencing apparatus was used for manual sequencing. After polymerization of gel, clamp, sealing tap and comb was removed. The top of gel was rinsed with 1x TBE to remove traces of unploymerized acrylamide. Plates were assembled in the apparatus and the comb was inserted in the top gap with its teeth towards the gel just to make leakproof well. The top chamber of apparatus was filled with 0.5x TBE and bottomchamber with 1x TBE buffer. The gel was pre-run at 65 watts for 45 min before loading the samples in order to attain the temperature of ~ 55^0C.

After the gel had attained temperature of ~55^0C, the wells were washed with 0.5 x TBE to remove traces of urea. The DNA samples were heated at 75^0C for 2 min on heating block just prior to loading. The wells were labelled with marker on plates (A, C, G, & T) and samples (3 µl each of ddA, ddC, ddG & ddT separately) were loaded in the wells. Stop solution was used as tracking dye. Second loading was done, when first loaded samples had reached half a way in the gel. Electrophoresis was stopped, after tracking dye had attained terminal position in the gel.

Autoradiography: After completion of electrophoresis, the plates were opened gently with spatula and one plate with gel was submerged in a large tray containing bandfixing solution (10% methanol and 10% acetic acid) for 30 min at RT. Then fixing solution was sucked off from tray and gel was taken on Whatman no.1 filter paper and was covered with saranwrap. This was dried on gel dryer at 80^0C for 2 hrs, exposed to Kodak film for 5-8 days and was developed using Kodak developer and fixer.

3.8 Sequence Analysis

3.8.1 BLAST (Basic Local Alignment Search TOOL) analysis

The sequences were analyzed using BLAST X program. BLAST X dynamically translates query and database sequence so the analysis is based on the polypeptide sequences, which is very reasonable due to existence of suitable degree of conservation among polypeptide sequences.

Web site: **http://ncbi.nlm.nih.gov/blast**

3.8.2 Multiple alignment

Multiple sequence alignment was done using program CLUSTAL W. The heuristic used in clustal w is based on phylogenetic analysis. In protein sequence alignment, different scoring matrices are used for each alignment based on expected evolutionary distance. Use of BLOSUM62 for close relationships and BLOSUM45 for more distant relationships are preferred, rather than the same scoring matrix for all alignments.

Web site: **http://www.ebi.ac.uk/clustalw**

3.9 Submission of Sequences to Database to get accession numbers

The sequences were submitted to EMBL database using its recent sequence submitting program WEBIN to get accession numbers (see list of accession numbers)

Web site: **http://www3.ebi.ac.uk/Services**

Chapter 4: Results

4.1 Isolation of reverse transcriptase conserved sequences from the chickpea genome

Putative retroelement reverse-transcriptase sequences (partial *pol* sequences) have been amplified utilizing the information that it is more or less conserved region of retrotransposons among all organisms with complex genomes. The conserved priming regions were selected to exclude all other retroelements, including other retrotransposon families. The selected regions are shown in fig: 9, both upstream and downstream primers designed were degenerate and conserved to corresponding regions. The downstream primer region chosen encodes the LYVDDMDP peptide motif, which is the most strongly conserved sequence in the reverse transcriptase protein reported by Voytas and Ausubel in 1988 and Grandbastien *et al* in1989. While the upstream primer region chosen encodes the DVKTAFLHG peptide motif, which is less conserved sequence in comparison to LYVDDMDP. The amplification of chickpea genomic DNA was done using these degenerate RTase (reverse transcriptase) primers (Fig: 9 and 10). The amplification of chickpea genomic DNA generated a large number of amplicons. The amplicons were ligated in pGEM T Easy vector and cloned in *E.coli* (DH5α strain). For plasmid miniprep, total 400 white colonies were selected and after restriction digestion of plasmids with EcoRI only 66 white colonies were found to have clones. Theoretically it is predicted that all white colonies should be clone positive, but the percentage of white bacterial colonies with partial pol sequence (reverse transcriptase) was around 12%. The clones from these sixty six white colonies were isolated and sequenced (Fig: 11). Altogether 43 clones out of the 66 were found to have characteristic homology with Tyl-*copia* reverse transcriptase gene sequences from data base (gene bank). While 7 clones clone no. CART 5, 6, 24, 81, 186, 204 and 273 showed no significant homology with *Tyl-copia* RTase gene sequences (see the list of Accession

Upstream primer (26-mer)

5′ GGGATCCA (T/C) (A/G) TC (A/G) TC (A/C/G/T)
AC (A/G) TA (C/A/G/T) A (A/G) (A/C/G/T) A **3′**

Downstream primer (27-mer)

5′ GGAATTCGA (T/C) GT (A/C/G/T) AA (A/G) AC
(A/C/G/T) GC (A/C/G/T) TT (T/C) (T/C) T **3′**

Fig: 9
Primers Designed from Conserved Regions of
Reverse Transcriptase

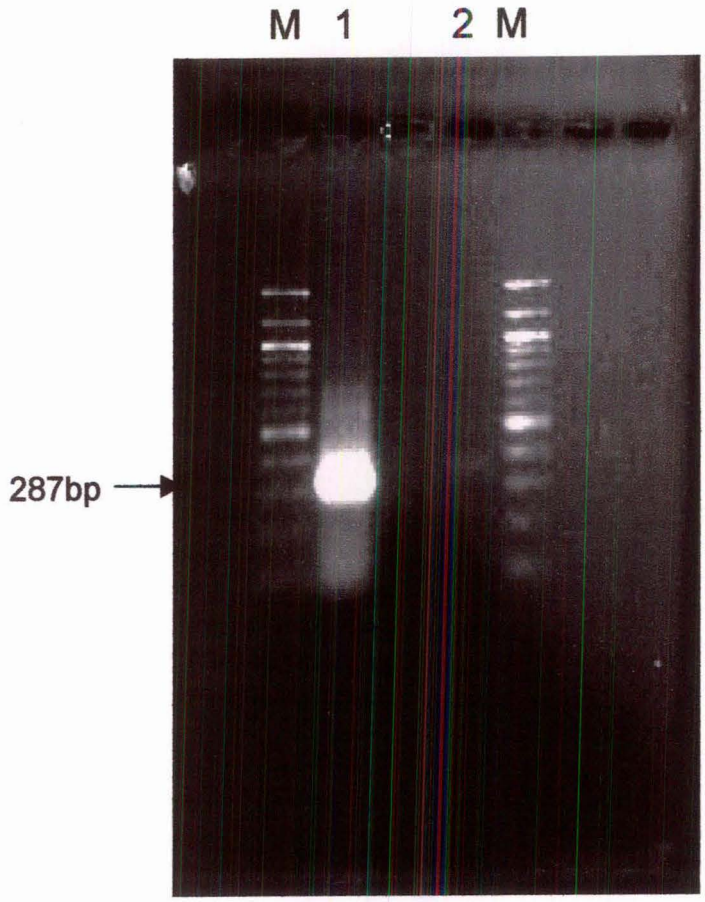


Fig: 10
Lane M-100 bp DNA ladder marker
Lane 1- PCR amplified product with 1.5mM $MgSO_4$
Lane 2- PCR amplified product with 2mM $MgSO_4$

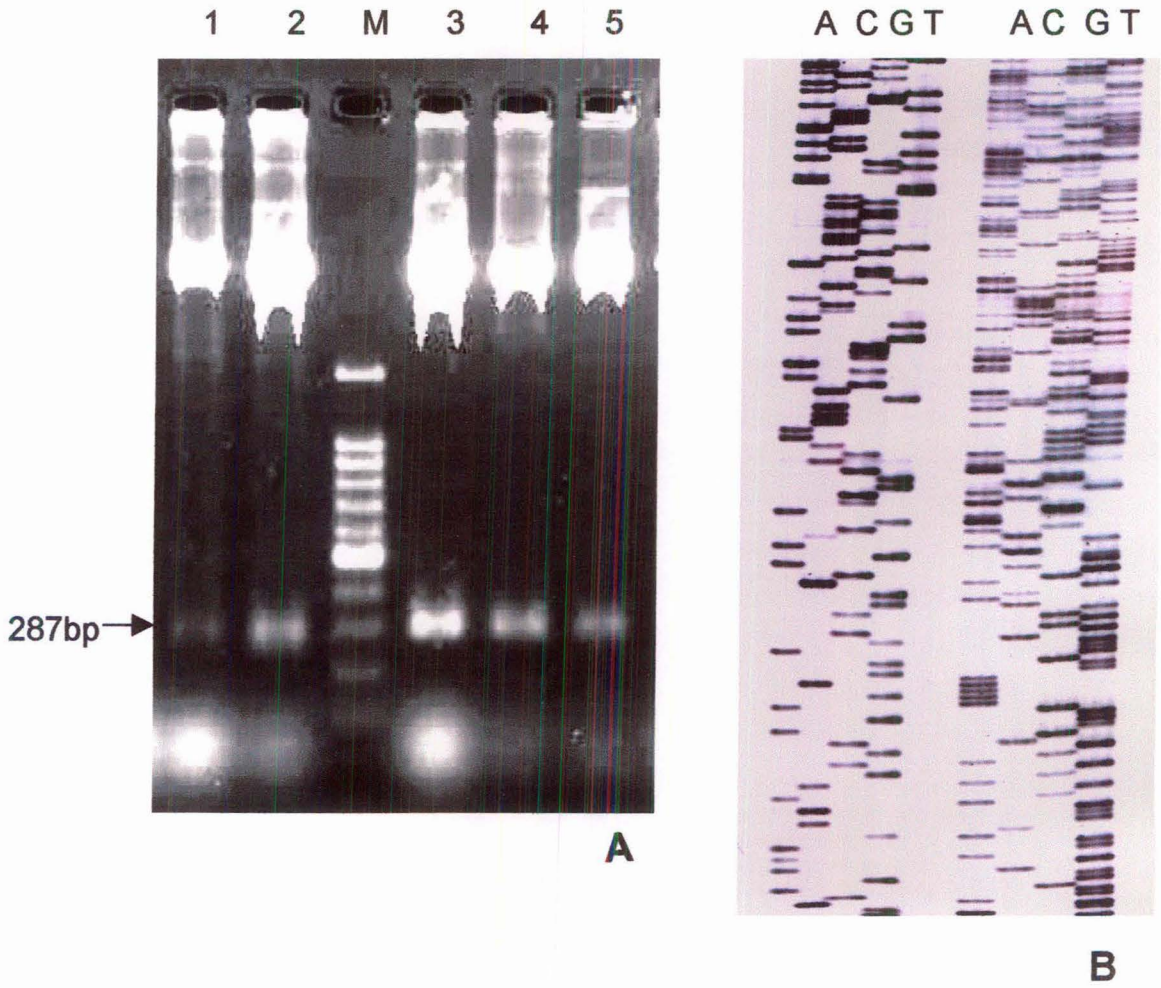


Fig: 11
(A) Plasmid digested with EcoRI
(B) Autoradiograph of Sequencing gel.

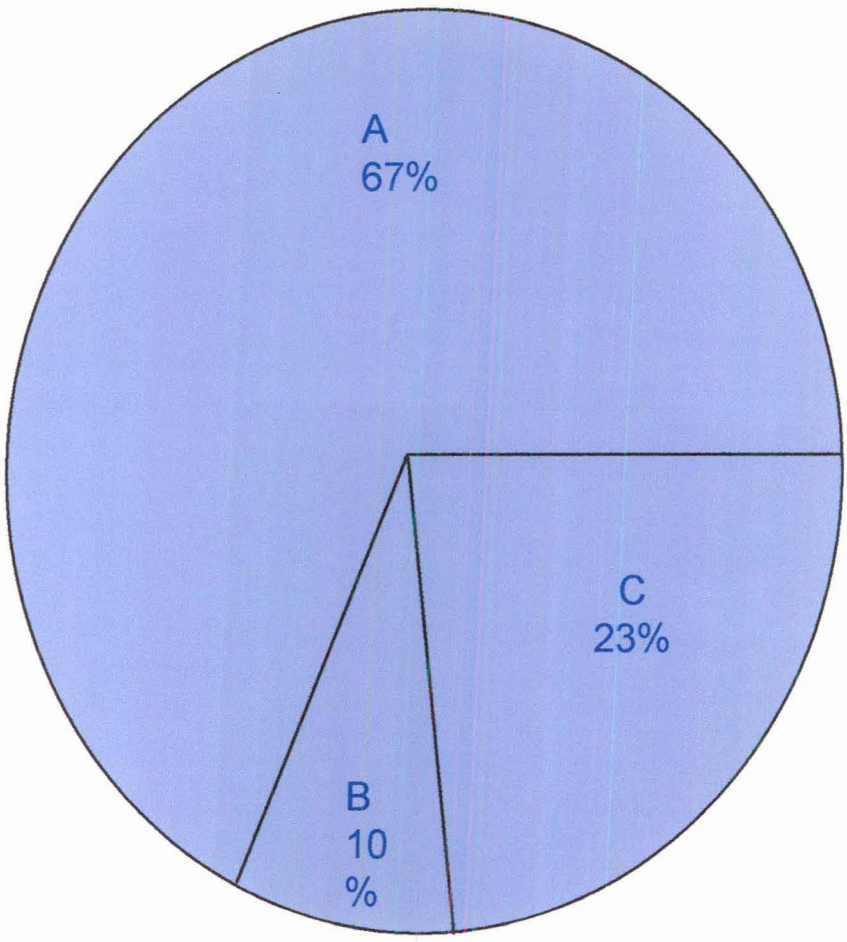


Figure: 12 Distribution of sequences in Chickpea based on our study
A = Reverse transcriptase sequences
B = Sequences with less homology to rt
C = Protein kinase like sequences
Scale - 360° = 100%

Alignment of RTase sequences

```
CART9     EFDVKTAFLHGDLE-EEIYME---QLEGFEVKV-TTSDTTKL-KLSLKHLKQAPR
CART73               DLE-EEIYRE---QLEGFEVKD-ELQLVCKL-KKSLYGVKQAPR
CART126              DLE-EEIYSE---VKTSFERIC-KEPLVCKL-KKSNHGLKQAPR
CART241              DLE-EEIYME---QLEGFESSD-TKQLRRKL-KKGLYGLKHAPR
CART67               DLESDHIYME---QPEGFKMFG-KEQLVCKL-KKSLQGLKQAPR
CART120              DLE-EEIYME---QPEGFECRR-KEQLVCKY-GKSLYEEKQAPR
CART103              DLE-DEIYMK---QPDGFLVKG-GSDDDARL-RKSLYGLKQAPR
CART345              DLE-EEIYMK---QPDGFLKVG-GEDYVCRL-RKSNCILKQAPR
CART77               DLE-EEIYMK---QPDGFLVKG-DASYVCRL-RKSLYGLKQAPR
CART105              DLE-CTIYMK---QPDGFLVKG-KEDYVCRK-RLSLYGLKQRPR
CART141              DLE-EKIYMK---QPDGFLVKG-CDDYVCRL-RKSLKLGRQAPR
CART13               DLE-EEIYMK---QPDGFLVKG-KMDVVCRL-RMSLYGLKQAPR
CART98               DLE-EEIYMK---QPDGVFLKG-KSSFHCRL-RKSLYGLKPAPP
CART155              DLE-EEIYMK---RPDGFLTAG-KEDVYCRL-RKSLYGSKQAPR
CART14               DLE-EEIYMK---QPDGFLVKG-KEDYVCRL-RKYSLGLKQAPR
CART40               DLE-EEIYMK---QGDLFLVKG-KEDYVCRL-RSKDYGLKQAPR
CART311              DLE-EEIYDN---QPDDFLFFG-KEDYVCRL-RKSLYGLKQAPR
CART389              DLE-EEIYMN---QPDGFLVKG-KEDYVCRL-RKSLYGLKQAPR
CART377              DLE-EEIYMK---QPDGFLVKG-KEDYVCRL-RKSLYGLKQAPR
CART135              DLE-EEIYMK---QPDGFLVKG-AEDYVCRL-RKSLYGLKQAAF
CART394              DLE-EEIYMK---QPDGFLVKG-KEDYVCRL-RKSLYGLKQAPR
CART33               DLE-EEIYMK---RPDGFSVKH-QEEYVCRL-RSSLSGLKQAPR
CART56               DLS-SEIYMK---QPDGFLKVM-DGDYVCRL-RKSLYGLFQTPR
CART55               DLE-KEIYMK---QPDGFLVQQ-KEDYVCRL-RSKLYGKKQPAR
CART364              DLE-EEIYMK---QLDGFLVVM-CEDYVCRL-RKSLYGVMQAQR
CART54               DLE-ELIYMK---RRDGFLVKG-KDDYVCRL-RKSLDVLKQAPR
CART324              DLE-KLTYMK---QPDFGLVKS-DGKYVCRL-RKSLYGLKQAPR
CART52               DLS-EEIYMK---QPDGFLEEF-LDEYVCRL-RKSLYHLKQIPR
CART97               DLL-EEIYMK---QPDGFLDEF-KEDYVCRL-RKSLYGLKQAPR
CART329              DLE-EEIYMK---QPDGKLMVS-KFDYVCRL-RKSLYGLKPSQR
CART123              DLH-GDIYMK---QPNGFLVKG-GKDYVRCL-IKSLYGLKRVPR
CART145              DLE-EEIYMK---QRNGFLVKG-DKKYVCRL-IKSLYGLKRVPR
CART210              DLE-EQIYMK---QPDGFLVFAGKEDCYVRL-RKSLYGLKSQPR
CART37               DLN-KTILMQ---QPYGFRIQG-KEDWVCLL-KRSVYGIKQSPR
CART157              DQN-ETILMQ---QPNGFRTQG-KEDWVCLL-KRSLYGLLESSR
CART50               DID-ETIYMV---QPENKSHGD-PKNMVCKL-RKSIYGCKQASR
CART366              DID-ETIVYM---QPENFMLGD-PKNMVCIL-RKSIYGLKQASR
CART387              DID-ETIYMV---QPENFVLGD-PKNMVCKL-RKSIYGLKQASR
CART61               DLE-EEIYMD---LPPGYSEHI-AANTVCKL-KKALYALKQSPR
CART384              DLE-EEIYMD---LPLGYSEHI-AANTVCKL-KKALYGLKQSPR
CART381              PGK--YVCSH-TYDTHNHQLYS-SQNQMREL-LDEIYGTIDEKL
CART10               AGN--SIGIR---RQNSFLGIHVVYVEVMMA-KQLIYRKKAALL
CART372              NRLLFYMVIW---RKQRVSKSK-VNSSLCAN-RKVCMGTSKHLD
```

Alignment of predicted peptide sequences of *Ty*1-*copia* reverse transcriptases of Chickpea.
CART- *Cicer arietimum* retro-transposon like element.


CART9 EFDVKT--AFLHGDLE-EEIYME--QLEGFEVKV--TTSDTTKLKLSL-KHL-KQAP
CART73 DLE-EEIYRE--QLEGFEVKD--ELQLVCKLKKSL-YGV-KQAP
CART126 DLE-EEIYSE--VKTSFERIC--KEPLVCKLKKSN-HGL-KQAP
CART241 DLE-EEIYME--QLEGFESSD--TKQLRRKLKKGL-YGL-KHAP
CART67 DLESDHIYME--QPEGFKMFG--KEQLVCKLKKSL-QGL-KQAP
CART120 DLE-EEIYME--QPEGFECRR--KEQLVCKYGKSL-YEE-KQAP
Petunia DLD-EEIYME--QPEGFEVKG--KENYVCRLKKSL-YGL-KQAP
Lycopersicon DLD-EEIYME--QPEGFEVKG--KENYICKLKKIL-YGL-KQAP
Solanum DLE-EEIYME--QPEGFKVEG--KENFVCKLKKSL-YGL-KQAP
Allium DLE-EEIYME--QPEGFKVPG--KEGLVCHLTKSL-YGL-KQAP
CART54 DLE-ELIYMK--RRDGFLVKG--KDDYVCRLRKSL-DVL-KQAP
CART324 DLE-KLTYMK--QPDFGLVKS--DGKYVCRLRKSL-YGL-KQAP
CART13 DLE-EEIYMK--QPDGFLVKG--KMDVVCRLRMSL-YGL-KQAP
CART98 DLE-EEIYMK--QPDGVFLKG--KSSFHCRLRKSL-YGL-KPAP
CART155 DLE-EEIYMK--RPDGFLTAG--KEDVYCRLRKSL-YGS-KQAP
CART14 DLE-EEIYMK--QPDGFLVKG--KEDYVCRLRKYS-LGL-KQAP
CART40 DLE-EEIYMK--QGDLFLVKG--KEDYVCRLRSKD-YGL-KQAP
CART377 DLE-EEIYMK--QPDGFLVKG--KEDYVCRLRKSL-YGL-KQAP
CART311 DLE-EEIYDN--QPDDFLFFG--KEDYVCRLRKSL-YGL-KQAP
CART389 DLE-EEIYMN--QPDGFLVKG--KEDYVCRLRKSL-YGL-KQAP
CART103 DLE-DEIYMK--QPDGFLVKG--GSDDDARLRKSL-YGL-KQAP
CART345 DLE-EEIYMK--QPDGFLKVG--GEDYVCRLRKSN-CIL-KQAP
CART77 DLE-EEIYMK--QPDGFLVKG--DASYVCRLRKSL-YGL-KQAP
CART105 DLE-CTIYMK--QPDGFLVKG--KEDYVCRKRLSL-YGL-KQRP
CART141 DLE-EKIYMK--QPDGFLVKG--CDDYVCRLRKSL-KLG-RQAP
CART135 DLE-EEIYMK--QPDGFLVKG--AEDYVCRLRKSL-YGL-KQAA
CART394 DLE-EEIYMK--QPDGFLVKG--KEDYVCRLRKSL-YGL-KQAP
CART33 DLE-EEIYMK--RPDGFSVKH--QEEYVCRLRSSL-SGL-KQAP
CART56 DLS-SEIYMK--QPDGFLKVM--DGDYVCRLRKSL-YGL-FQTP
CART55 DLE-KEIYMK--QPDGFLVQQ--KEDYVCRLRSKL-YGK-KQPA
CART52 DLS-EEIYMK--QPDGFLEEF--LDEYVCRLRKSL-YHL-KQIP
CART97 DLL-EEIYMK--QPDGFLDEF--KEDYVCRLRKSL-YGL-KQAP
CART329 DLE-EEIYMK--QPDGKLMVS--KFDYVCRLRKSL-YGL-KPSQ
CART364 DLE-EEIYMK--QLDGFLVVM--CEDYVCRLRKSL-YGV-MQAQ
CART123 DLH-GDIYMK--QPNGFLVKG--GKDYVRCLIKSL-YGL-KRVP
CART145 DLE-EEIYMK--QRNGFLVKG--DKKYVCRLIKSL-YGL-KRVP
CART210 DLE-EQIYMK--QPDGFLVFAG-KEDCYVRLRKSL-YGL-KSQP
CART37 DLN-KTILMQ--QPYGFRIQG--KEDWVCLLKRSV-YGI-KQSP
CART157 DQN-ETILMQ--QPNGFRTQG--KEDWVCLLKRSL-YGL-LESS
CART61 DLE-EEIYMD--LPPGYSEHI--AANTVCKLKKAL-YAL-KQSP
CART384 DLE-EEIYMD--LPLGYSEHI--AANTVCKLKKAL-YGL-KQSP
CART50 DID-ETIYMV--QPENKSHGD--PKNMVCKLRKSI-YGC-KQAS
CART387 DID-ETIYMV--QPENFVLGD--PKNMVCKLRKSI-YGL-KQAS
CART366 DID-ETIVYM--QPENFMLGD--PKNMVCILRKSI-YGL-KQAS
Tobacco TRNCGVIPFR-MGLEGEICGVH-PLNLRKKFSSLLQFSI-AATT
CART10 AGN--SIGIR--RQNSFLGIH--VVYVEVMMAKQLIYRK-KAAL
CART381 GPG-KYVCSH--TYDTHNHQLYSSQNQMRELLDEI-YGT-IDEK
CART372 NRL--LFYMVIWRKQRVSKSK--VNSSLCAN-RKVCMGT----S

Alignment of predicted peptide sequences of *Ty*1-*copia* reverse transcriptases of Chickpea (*Cicer arietinum*) with predicted peptide *Ty*1-copia reverse transcriptase sequences *Petunia, Allium, Lycopersicon, Solanum* and *Tobacco Tnt*1
CART- *Cicer arietinum* retro-transposon like element.

numbers; Fig: 12). Rest of clones were identified as partial protein kinase gene sequences with a variable degree of homology.

4.2 Multiple *Ty*1-*copia* group sequences in chickpea (*Cicer arietinum*)

The multiplicity of *Ty*l-*copia* group sequences in chickpea by PCR analysis was revealed by using degenerate pair of primers. The translated BLAST X analysis of novel 43 clones gave very surprisingly inference, that no two sequences were identical (see alignment and phylogenetic tree, Fig: 13a, b, c, d), though all obviously had significant homology to the reverse transcriptase region of *Ty*l-*copia* retrotransposons of potato, tomato, onion, tobacco and petunia (Fig: 14a, b, c).

The sequence similarities among the chickpea clones were evaluated by converting the nucleotide data into peptide sequences and compared each clone sequence with the all clones altogether, using CLUSTAL W and CLUSTAL X multiple alignment computer programs. The entire set of clones was compared with corresponding regions of known retrotransposons in this analysis. All of the 43 sequences examined by us were obviously derived from *Ty*l-*copia* group retro-transposons as confirmed by their sequence similarities with RTase of different *Ty*1-*copia* of various plants including *Tnt*l of tobacco (Grandbastein et al., 1989). The sequences do not show strong homology to reverse transcriptase domain of retroviruses (Xiong and Eickbush 1990) i.e all reverse transcriptase sequences are specifically from *Ty*1-*copia* group of retrotransposons. They even do not show any significant homology to their closest group *gypsy* retrotransposons. This is due to the fact that *gypsy* group of retrotransposons are phylogenetically more near to the retroviruses than *copia* group of retrotransposons. Also in *Drosophila gypsy* retrotransposons are shown to have infectious activity like retroviruses upon manipulation of its *env* coding region, but in case of *copia* group no such activity is evident till date. However none of two out of the 43 chickpea derived reverse transcriptase sequences aligned are identical. This shows that PCR band might had more than 43 different sequences which would be having characteristics homology to the reverse

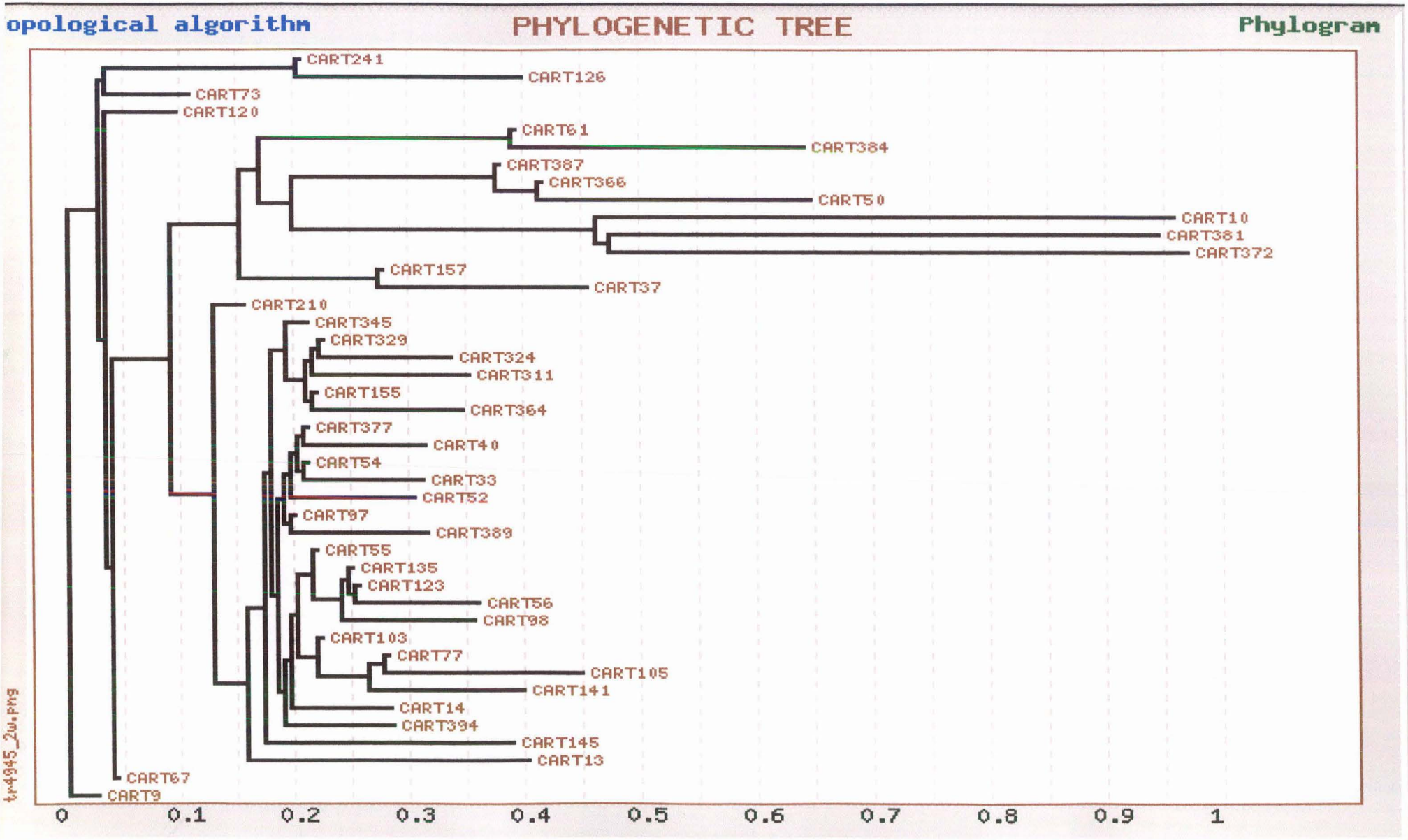


Fig: 13a Phylogenetic tree of predicted peptide sequences of Ty1-copia reverse transcriptases of chickpea. Divergences in distance units are indicated by branch lengths. CART- *Cicer arietinum* retrotransposon like element.

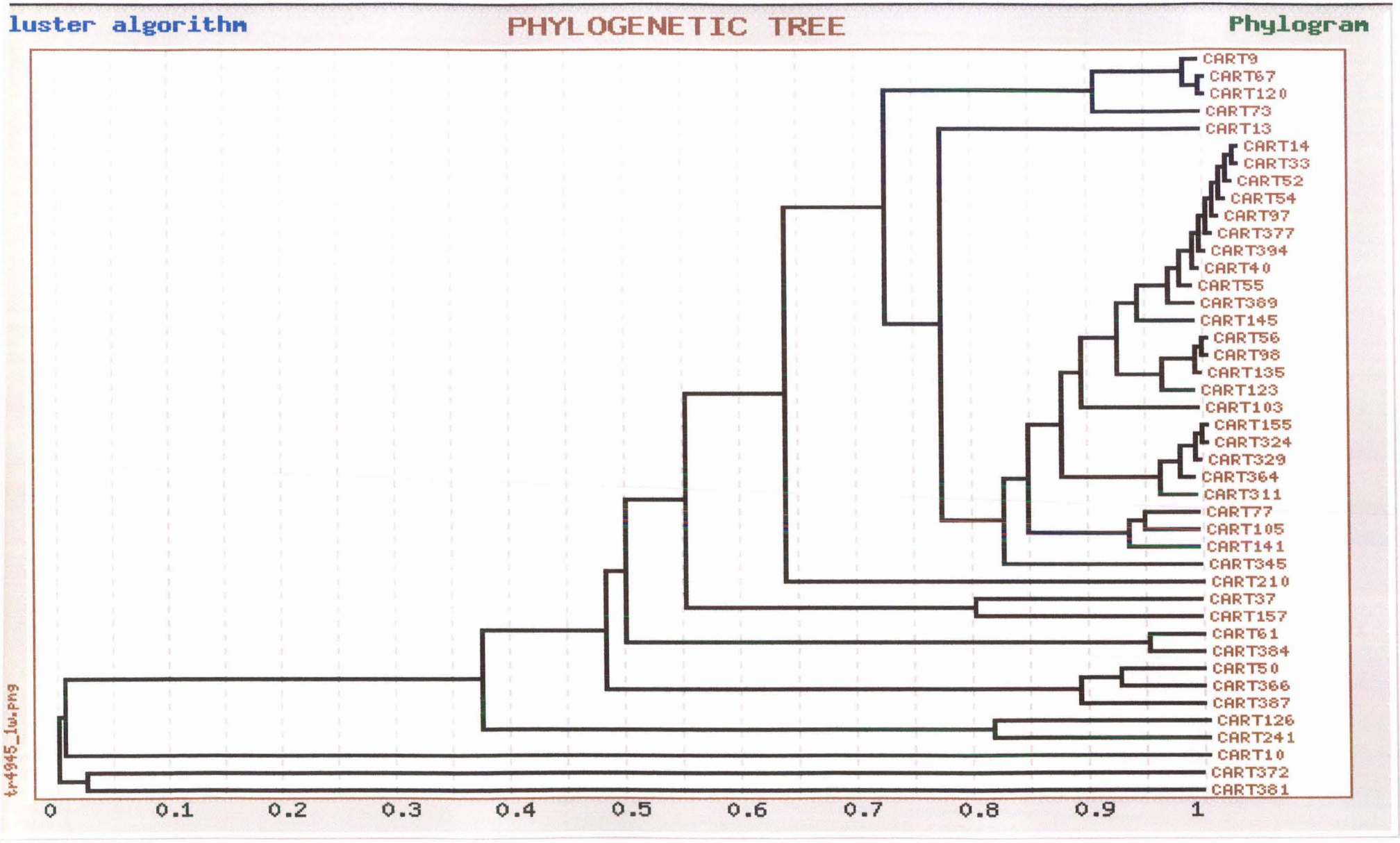


Fig: 13b Phylogenetic tree of predicted peptide sequences of Ty1-copia reverse transcriptases of chickpea. Divergences in distance units are indicated by branch lengths. CART- *Cicer arietinum* retrotransposon like element.

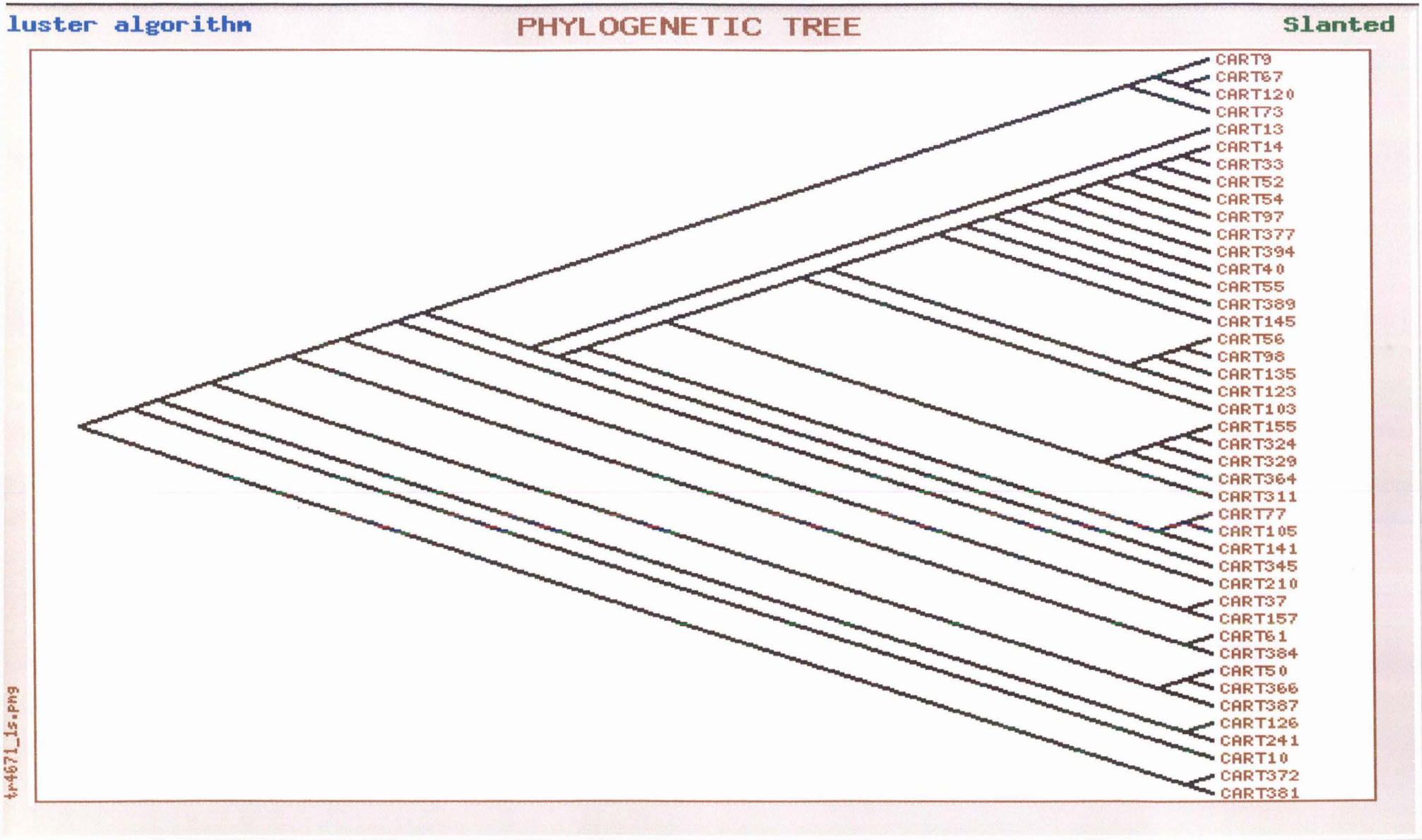


Fig: 13c Phylogenetic tree of predicted peptide sequences of Ty1-copia reverse transcriptases of chickpea. Divergences in distance units are indicated by branch lengths. CART- *Cicer arietinum* retrotransposon like element.

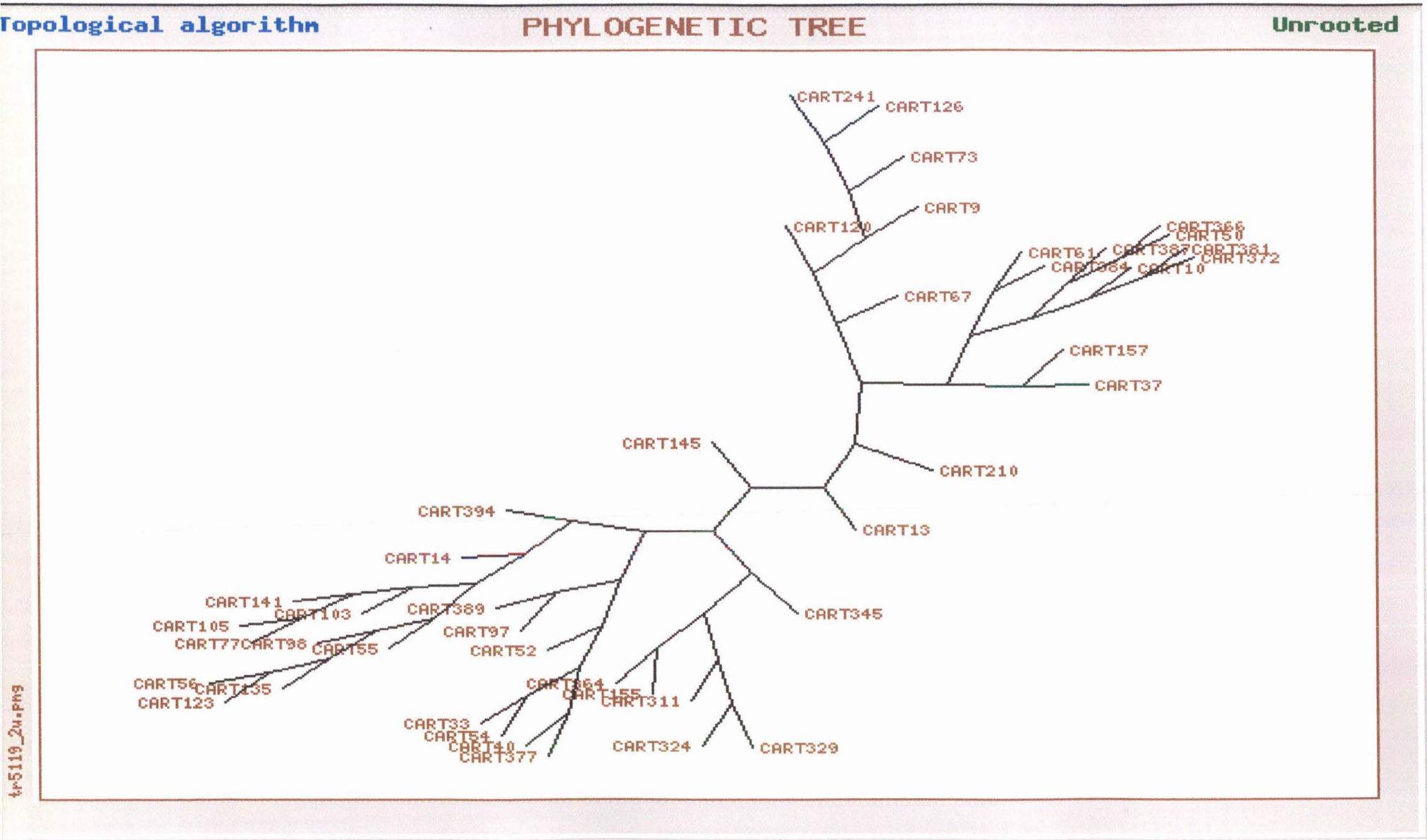


Fig: 13d Phylogenetic tree of predicted peptide sequences of Ty1-copia reverse transcriptases of chickpea. Divergences in distance units are indicated by branch lengths. CART- *Cicer arietinum* retrotransposon like element.

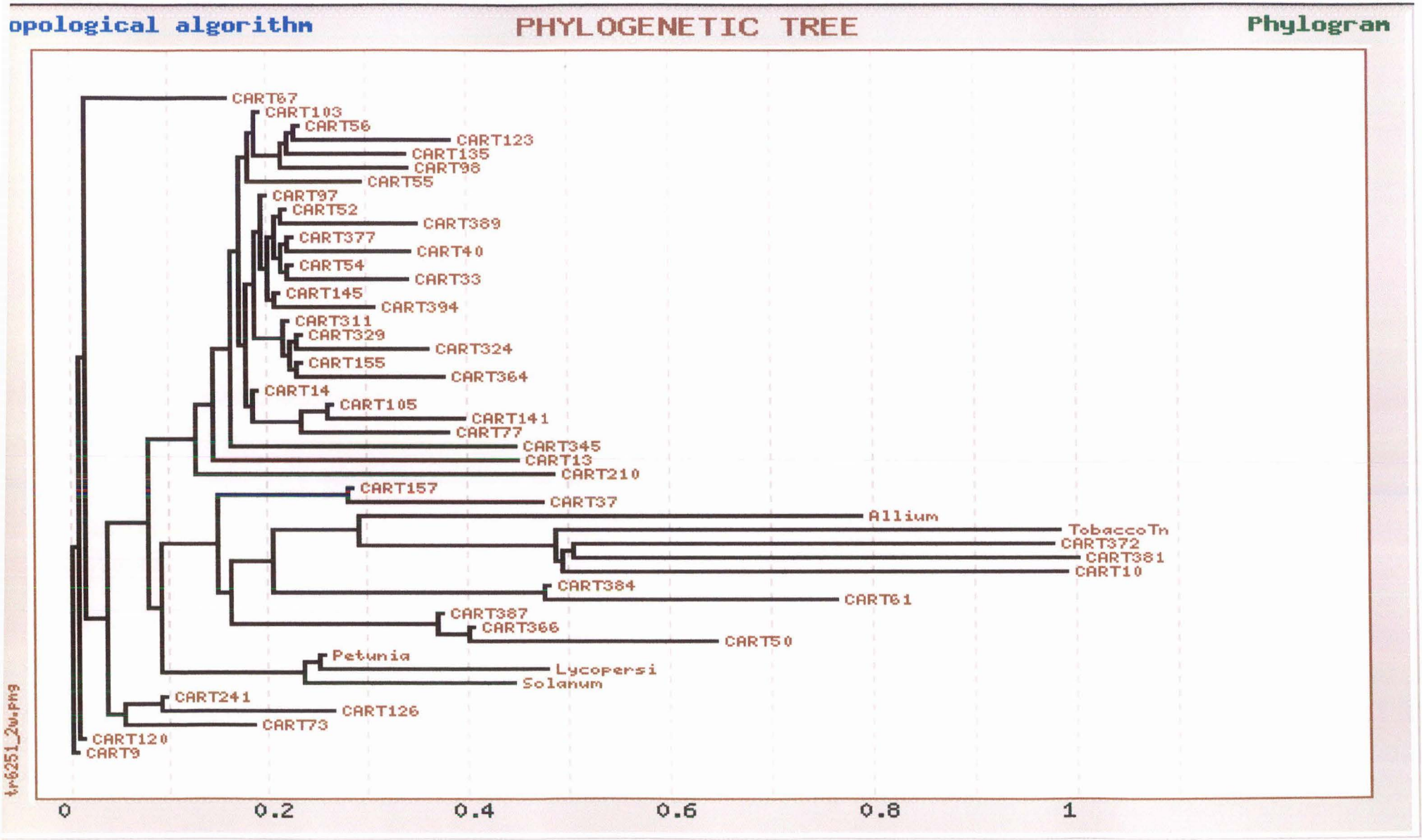


Fig: 14a Phylogenetic tree of predicted peptide sequences of Ty1-copia reverse transcriptases of chickpea, *Petunia, Allium, Lycopersicon, Solanum* and *Nicotiana* . Divergences in distance units are indicated by branch lengths.
CART- *Cicer arietinum* retro-transposon like element.

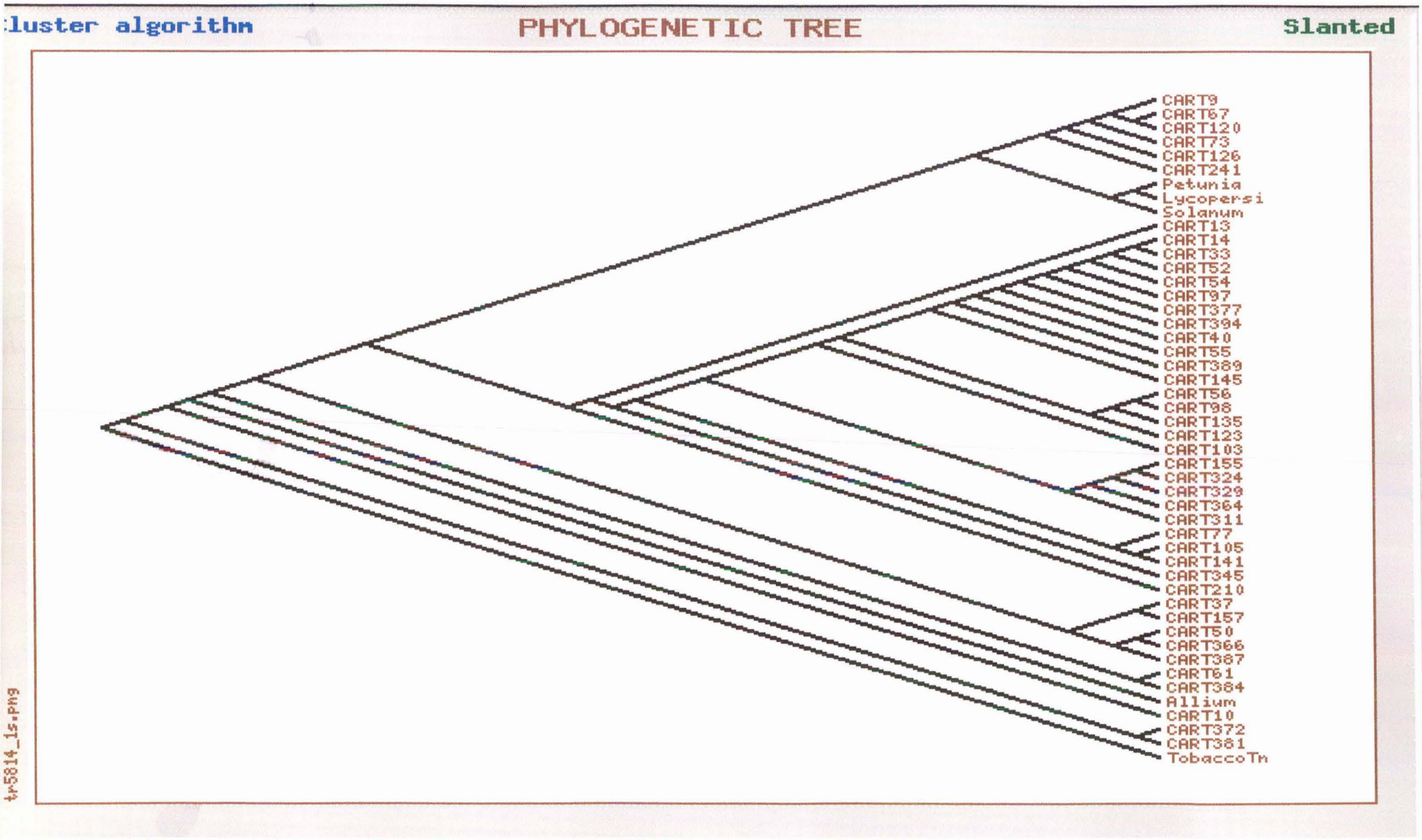


Fig: 14b Phylogenetic tree of predicted peptide sequences of Ty1-copia reverse transcriptases of chickpea, *Petunia, Allium, Lycopersicon, Solanum* and *Nicotiana* . Divergences in distance units are indicated by branch lengths.
CART- *Cicer arietinum* retro-transposon like element.

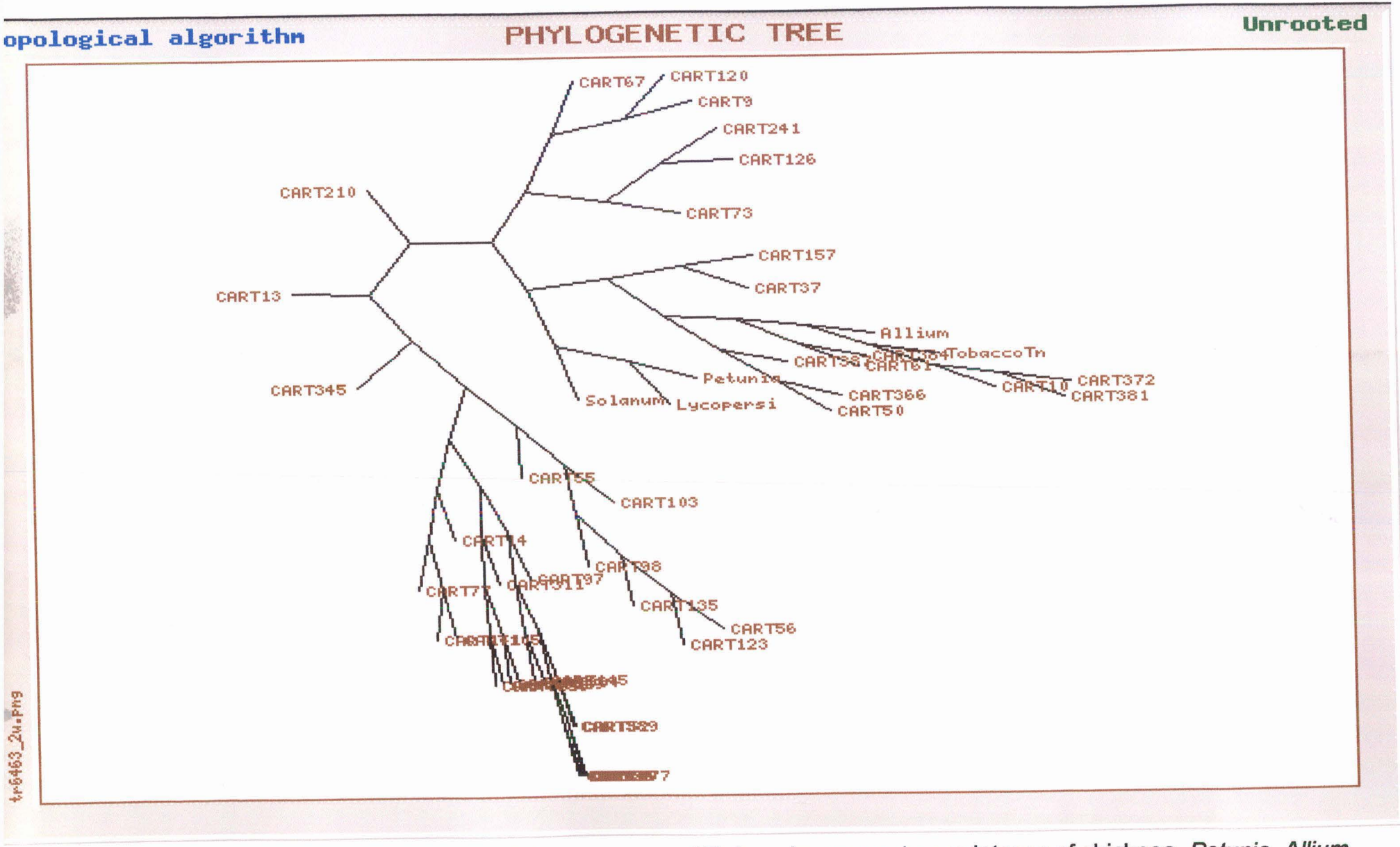


Fig: 14c Phylogenetic tree of predicted peptide sequences of Ty1-copia reverse transcriptases of chickpea, *Petunia, Allium, Lycopersicon, Solanum* and *Nicotiana* . Divergences in distance units are indicated by branch lengths.
CART- *Cicer arietinum* retro-transposon like element.

transcriptase region of *Ty*l-*copia* group retrotransposons because we picked up clones randomly and there might be chances of escaping some.

4.3 Classification of reverse transcriptase sequences

Though the sequences are not identical but some of these chickpea RT sequences are similar to each other. The phylogenetic tree was drawn to reveal their degree of homology among themselves and to categorize them in different sub-groups. Based on their phylogenetic relationships all 43 chickpea RT sequences fall into 9 sub-groups, which are represented by one of their representative members from each sub-group (Feng and Doolittle, 1987; Fig: 15a, b, c). The apparent degree of divergence, in their amino-acid sequences between the members of different *Ty*l-*copia* sub-groups is around 20-55% but it is 4-20% among the members of same *Ty*l-copia subgroup.

All of these putative partial *pol* (reverse transcriptase) sequences were aligned using CLUSTAL W computer program and phylogenetic trees were constructed to compare them with each other and with the reverse transcriptase sequences of *copia* group retrotransposons of other plants. Based on alignment and phylogenetic trees the following results were interpreted:

All the forty three chickpea *Ty*1-*copia* retrotransposon like sequences could be divided into nine sub-groups based on their amino acid divergence and origin. For origin relationships the phylogenetic tree constructed was considered as major key factor, thus the sequences branched from nearest single point of a major branch were collectively combined in the same sub-group (Fig: 13c, 15b). The percentage amino acid homology was also taken in account while classifying these sequences. In this way the 9 sub-groups were formed and each named after the top sequence of the sub-group (Table: 2). The sub-grouping was very clear from slanted phylogenetic tree (Fig: 15b). The first sub-group was named CART9 represented three other sequences CART67, CART120 and CART73 including itself in its sub-group. Similarly second sub-group was represented by CART13 with a total of twenty six sequences CART14, 33, 52, 54, 97, 377, 394, 40, 55, 389, 145, 56, 98,135, 123, 103, 155,

Table: 2

Classification of chickpea *Ty*1-*copia* retrotransposon like sequences on the basis of amino acid homology

Sub-group S. no.	Name of sub-group	No. of members	Members	% Intra sub-group range of homology	% Inter sub-group range of homology
1	CART9	4	CART9, 67, 120, 73	91-96	53-74
2	CART13	26	CART13, 14, 33, 52, 54, 97, 377, 394, 40, 55, 389, 145, 56, 98, 135, 123, 103, 155, 324, 329, 364, 311, 77, 105, 141, 345	82-90 (76% only for CART345)	50-78
3	CART210	1	CART210		44-57
4	CART37	2	CART37, 157	81	45-52
5	CART61	2	CART61, 384	94	42-49
6	CART50	3	CART50, 366, 387	96	38-51
7	CART126	2	CART126, 241	80	61-69
8	CART10	1	CART10		Un-significant, but good with gene bank seq.
9	CART372	2	CART372, 381	90	52-57

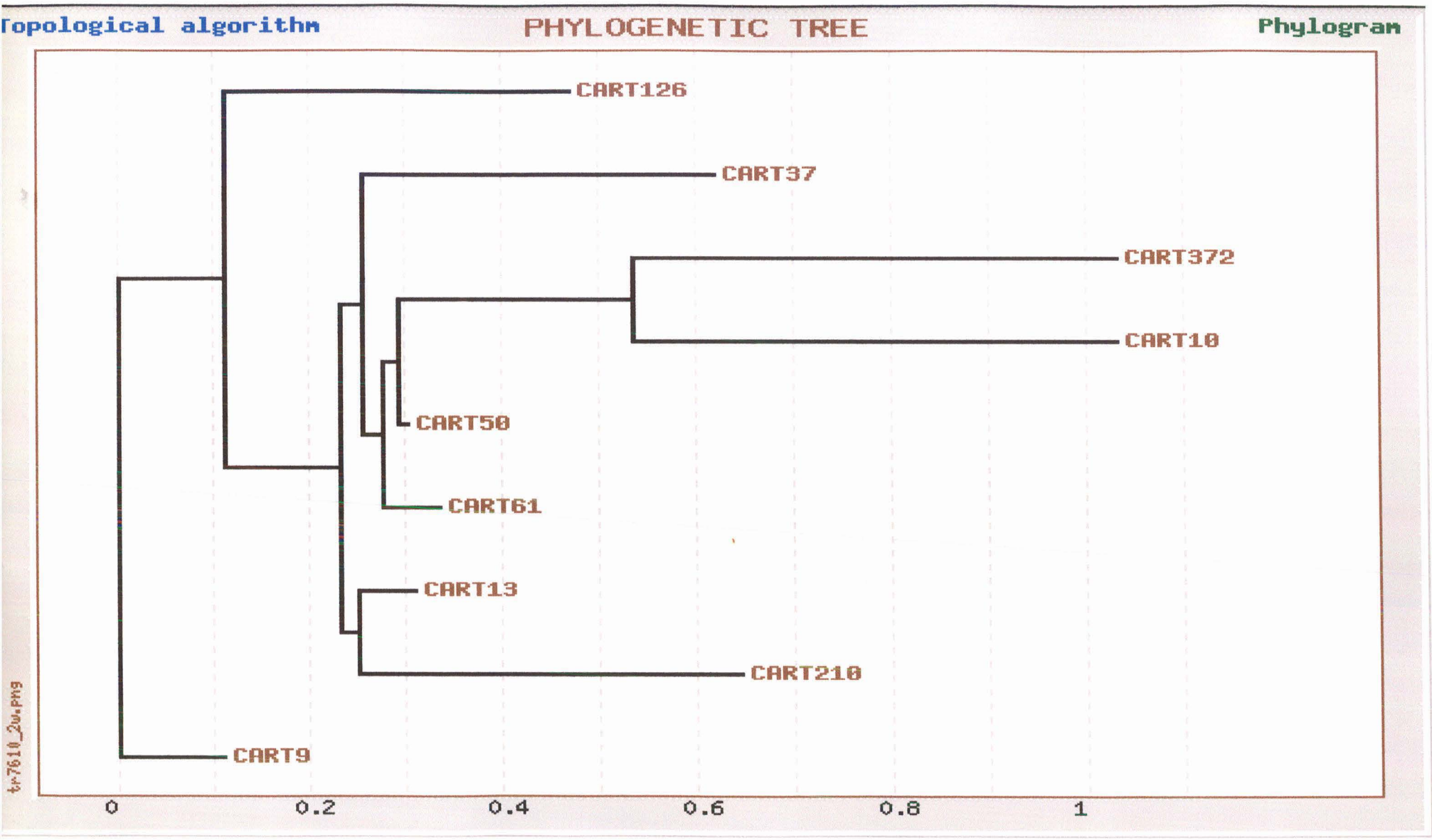


Fig: 15a Phylogenetic tree of predicted peptide sequences of selected sub-group members of Ty1-*copia* reverse transcriptases of chickpea. Divergences in distance units are indicated by branch lengths. CART- *Cicer arietinum* retro-transposon like element.

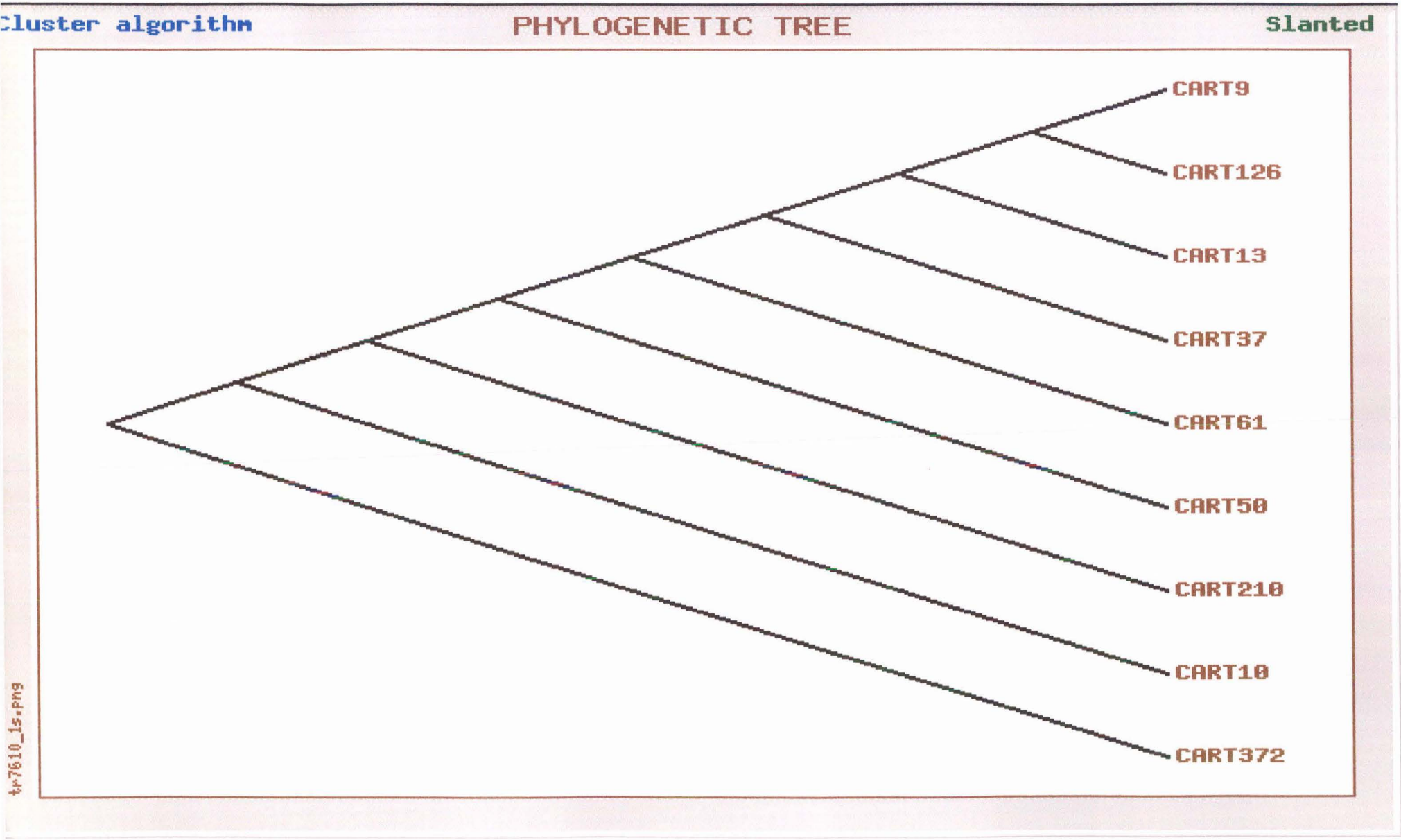


Fig: 15b Phylogenetic tree of predicted peptide sequences of selected sub-group members of Ty1-*copia* reverse transcriptases of chickpea. Divergences in distance units are indicated by branch lengths. CART- *Cicer arietinum* retro-transposon like element.

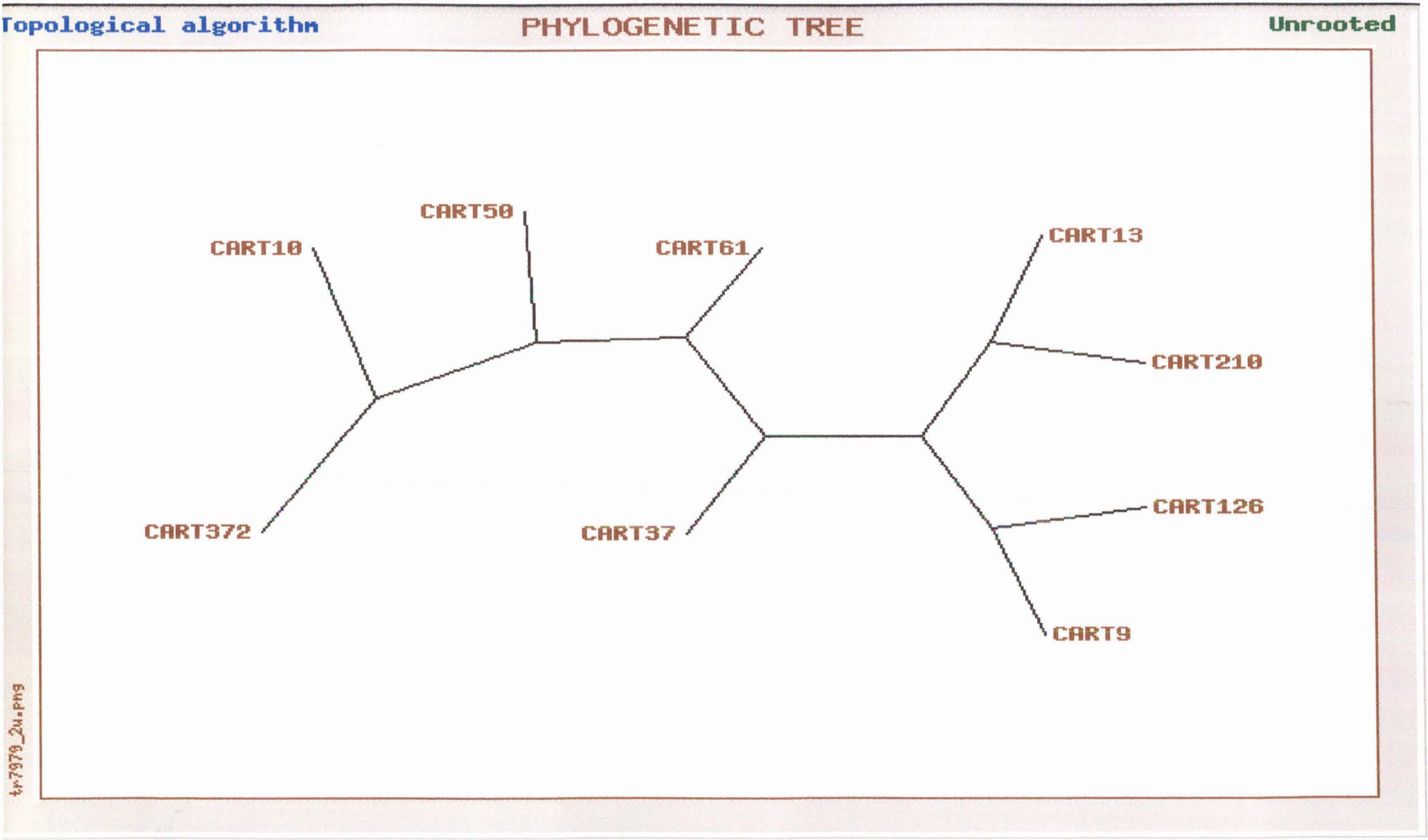


Fig: 15c Phylogenetic tree of predicted peptide sequences of selected sub-group members of Ty1-*copia* reverse transcriptases of chickpea. Divergences in distance units are indicated by branch lengths. CART- *Cicer arietinum* retro-transposon like element.

324, 329, 364, 311, 77, 105, 141 and CART345. It was the largest sub-group with all sequences from a common origin. Sub-groups 3rd and 8th were named after their lone members CART210 and CART10 respectively. Two member sub-groups, sub-group no. 4, 5 and 9 were represented by CART37 (another member CART157), CART61 (another member CART384) and CART372 (another member CART381 respectively, while CART50 represented itself and other two members CART387 and CART366 in sixth sub-group.

The amino acid divergence among the members of sub-group CART9 was found in the range of 4-9%, the minimum divergence was between CART9 and CART120 which was 4%. Similar results within the same sub-group were shown by sub-group CART61 and CART50. But in case of sub-group CART13 intra sub-group amino acid divergence was in the range of 10-18%. In spite of its large size with 26 members the amino acid divergence range was fairly small, the only member CART345 showed 24% amino acid divergence. The sequences of sub-group CART37 and sub-group CART126 showed 19% and 20% intra sub-group amino acid divergence respectively. Thus overall intra sub-group amino acid divergence was in a fair range of 4-20% in our study of chickpea *Ty*1-*copia* retrotransposon like sequences. The representative member of sub-group named CART372 itself showed unexpected large amino acid divergence when aligned with CART381 of the same sub-group. This could be explained on the basis of their phylogeny. Although they had branched from a common point but both diversified independently during the course of evolution.

The inter sub-group alignment comparison showed overall 20-55% amino acid divergence, sub-group named CART9 was 26-47% amino acid divergent to all other sub-groups except sub-group CART126, for which it was only 15% divergent due to their together diversification. This sub-group also showed 60-70% amino acid homology to reverse transcriptase sequences of *Petunia, Allium, Lycopersicon* and *Solanum tuberosum*. CART13 represented sub-group was found to be more closely related to the partial *pol* sequences of *Petunia, Allium, Lycopersicon* and *Solanum tuberosum* than to other sub-group

RT sequences of chickpea, because for gene bank sequences amino acid divergence was 22-34%, for chickpea 28-50%

The apparent inter sub-group amino acid divergence for the rest of sub-group was in range of 43-55% except CART126 represented sub-group. The clone CART126 showed 58% amino acid divergence when aligned with CART61.

Alignment and translated BLAST X analysis of all sub-groups showed significant homologies of 46-76% to reverse transcriptase region of *Ty*1-*copia* retrotransposons of *Petunia, Allium, Lycopersicon* and *Solanum* etc. This was also clearly supported by the positions of sequences in rooted and un-rooted phylogenetic trees (Fig: 14a, b, c).

Phylogenetic trees were also constructed by taking representative member of each sub-group and it was clear from these that sub-grouping was perfectly done based on their intra and inter-sub-group amino acid divergence. This was also beautifully supported by slanted phylogenetic tree showing uniform branching of sub-group representatives (Fig: 15b).

In addition sub-group 5 (CART61) and subgroup 7 (CART126) have more than 48% degree of divergence between their amino-acid sequences i.e 52% homology. Also two member CART372 and CART381 of same sub-group have ~ 40% degree of divergence but they are grouped together because the diversification is expected to arise in very later stages of evolution not at the time of the origin of these sequences from same sequence source. These data show that there exists a very high degree of sequence heterogeneity among *Ty*l-*copia* group members in the chickpea genome, with respect to both the number of different types of retrotransposon sequences and the degree of sequence divergence among many of these fragments. For instance, 80% amino acid sequence heterogeneity is observed between the least similar clones CART10 and CART241, which is in contrast to the grouping of different families of retrotransposons on the basis of their homology and heterogeneity because this degree of divergence between members of same group *Ty*l-*copia* exceeds the degree of divergence between the *copia* and 1731 retrotransponson

of *Drosophila* which is 59%. The degree of divergence between most similar sequences CART9 and CART120 observed is 4% which is again found in contrast to the degree of divergence found between different copies of the *copia* retrotransponsons of *Drosophila*, (generally less than 2% predicted amino acid divergence between individual copies (Mount and Rubin 1985; Emori et al., 1985).

The unexpected heterogeneity data would put a question mark on our results. So, in order to prove reliability of our experiment we analyzed PCR control run without added genomic DNA fragment to check the possibility that, are these sequences derived from cross-contamination with exogenous DNA and are not actually present in chickpea genome. But our data is found to be enough satisfactory and reliable because no amplified fragments were found in PCR control. Another loop-hole would point out that this heterogeneity results are due to PCR artifacts. So, to avoid this possibility, using the same degenerate RTase oligonucleotide primers, we amplified 10^6- fold dilutions of two different chickpea PCR clones. When the sequences of clones were compared to their corresponding sub clones, there was no case of any significant differences with in any pair. In our lab using same RTase primers *panzee Tyl-copia* retrotransposon has been isolated from *Pigeon pea* (Lall *et al.,* 2002). Therefore, our experimental approach does not, by itself, generate the observed heterogeneity.

From these observations we can also assume that there exists a very large number of retrotransposons which gives complexity to chickpea genome and contributes to largeness of genome. During comparison of these RT sequences with the RT sequences of *Tyl-copia* of other plants increases the complexity of branching in the existing phylogenetic tree i.e., when only chickpea RT sequences are aligned and phylogenetic tree is drawn to show their evolutionary relationships among themselves, the branching in tree is less complex and not very far from their rooting points. The comparison of representative members of corresponding sub-groups with RTase amino acid sequences from database shows that there is no increase in branch lengths of

```
CART77 GIRCQNCFLHGDLEEEIYMKQPDGFLVKGKEDYVCRLRKSLYGLKQAPRQWYKKFESVMC
Tto1   ------AFLHGDLEEEIYMEQPDGFQQKGKEDYVCRLRKSLYGLKQAPRQWYKKFESVMG
          .************:*****  ********************************

CART77           EQGYRKTTSDHCVFVKKFADDDFLILLL
Tto1             QHGYKKTTSDHCVFAQKFSDDDFIILLL
                 ::**:*********.:**:****:****
```

Alignment of predicted peptide sequences of CART77 and reverse transcriptase of *Tto*1 of Tobacco.
CART- *Cicer arietinum* retro-transposon like element.

phylogenetic tree (Fig: 15a). Also, the diversity of *Tyl-copia* RTase sequences within chickpea genome is comparable to that between the potato and tobacco or between any other interspecies genomes.

4.4 Predicted relation of reverse transcriptase sequences to functional retrotransposons

About 55% of the chickpea retrotransposon sequences seem to be from non-functional retrotransposons due to presence more than one translational stop codons in all six reading frames. Twenty four RT sequences out of 43 do not bear translational stop codons in at least one reading frame, so these are expected to be partial sequences of functional retrotransposons. The most interesting finding is 85.36% homology of clone CART 77 (Acc No. AJ535867) to *Tto*1 RTase amino acid sequences and *Tto*1 is functional retrotransposon in *N. tabaccum* (See alignment of CART77 and *Tto*1). This homology exceed the inter-sub-group homology among chickpea RTase sequences and was in range of intra sub-group homology, which is very interesting being from different plant species. Eighty two amino acid long RTase of tobacco *Tto*1, when aligned with chickpea CART 77 (Accession no. AJ535867) amino acid sequence. It is found that out of 89 amino acids, 70 of their amino acids are identical, 7 conserved and 2 semi-conserved. Thus these features of CART 77 in its first 5′-3′ reading frame contribute towards its greater degree of being functional.

Later on some best sequences from these RTase sequences will be used as molecular probes to fish-out complete retrotransposon.

4.5 Stress induced expression of retrotransposons

We also have started work to see stress induced retrotransposon expression. These stresses may be biotic and/or abiotic including UV (Ultraviolet light), drought and salicylic acid etc. Five transcriptionally active RTase sequences have been isolated and submitted to database (named as CARE 2, 3, 4, 5 and 7; see the list of accession numbers enclosed). These all

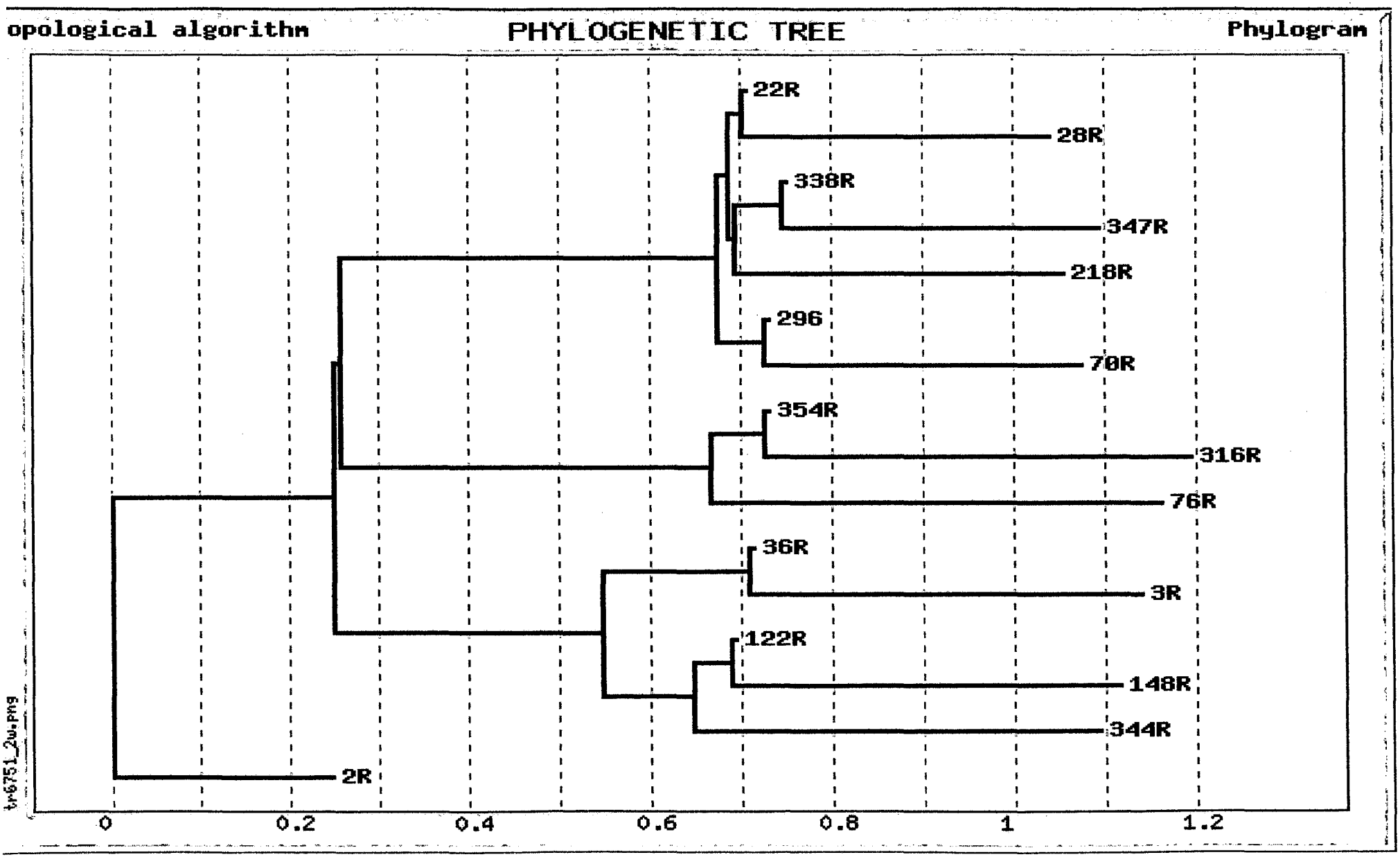


Fig: 16a Phylogenetic tree of predicted peptide sequences of partial protein kinase like sequences of chickpea. Divergences in distance units are indicated by branch lengths. R-Terminology

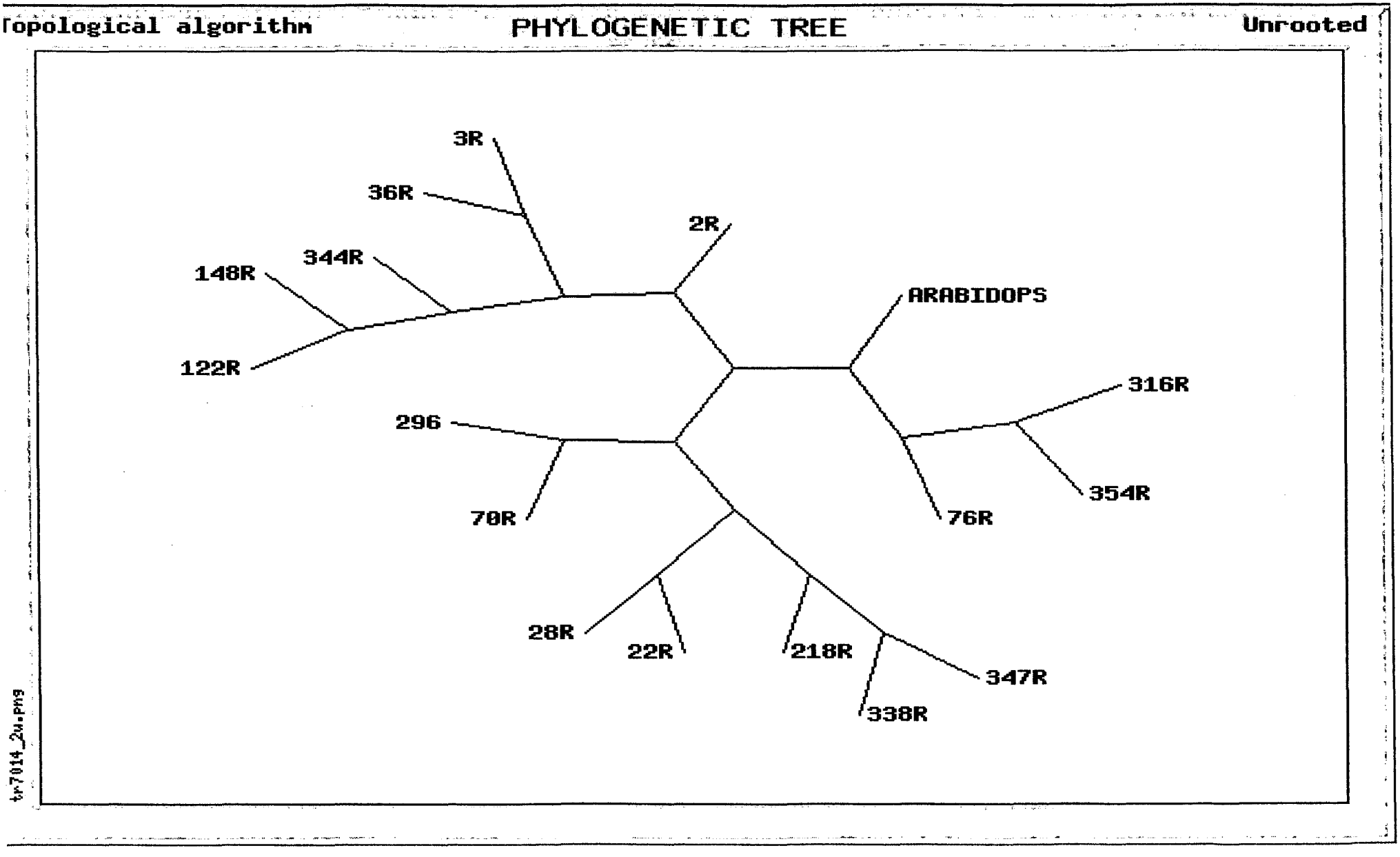


Fig: 16b Phylogenetic tree of predicted peptide sequences of partial protein kinase like sequences of chickpea. Divergences in distance units are indicated by branch lengths. R-Terminology

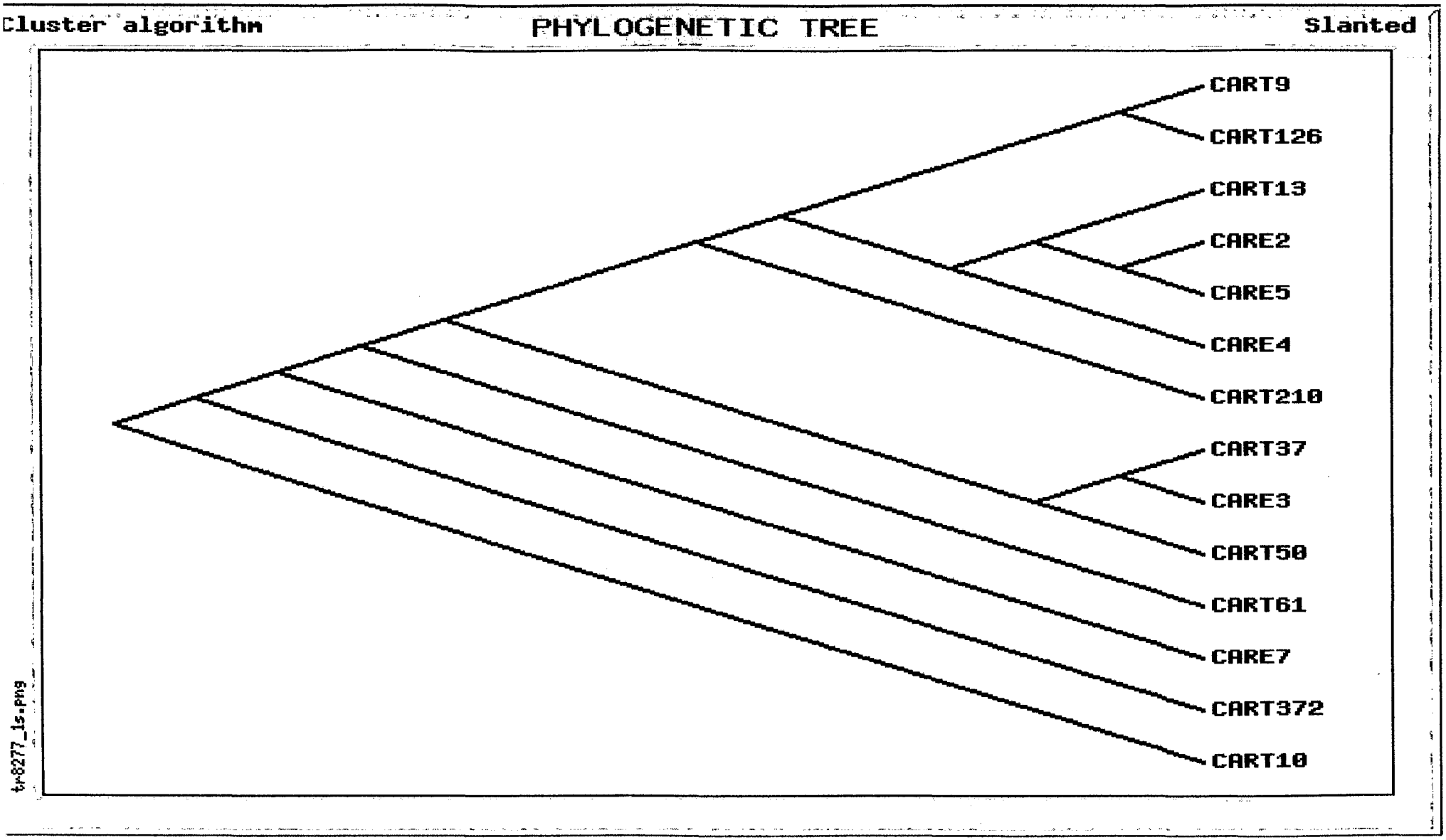


Fig: 17a Phylogenetic tree of predicted peptide sequences of Ty1-copia retrotransposon like sequences of chickpea. Divergences in distance units are indicated by branch lengths. CART- *Cicer arietinum* retrotransposon like sequences, CARE- Transcriptionally active *Cicer arietinum* retroelement like sequences.

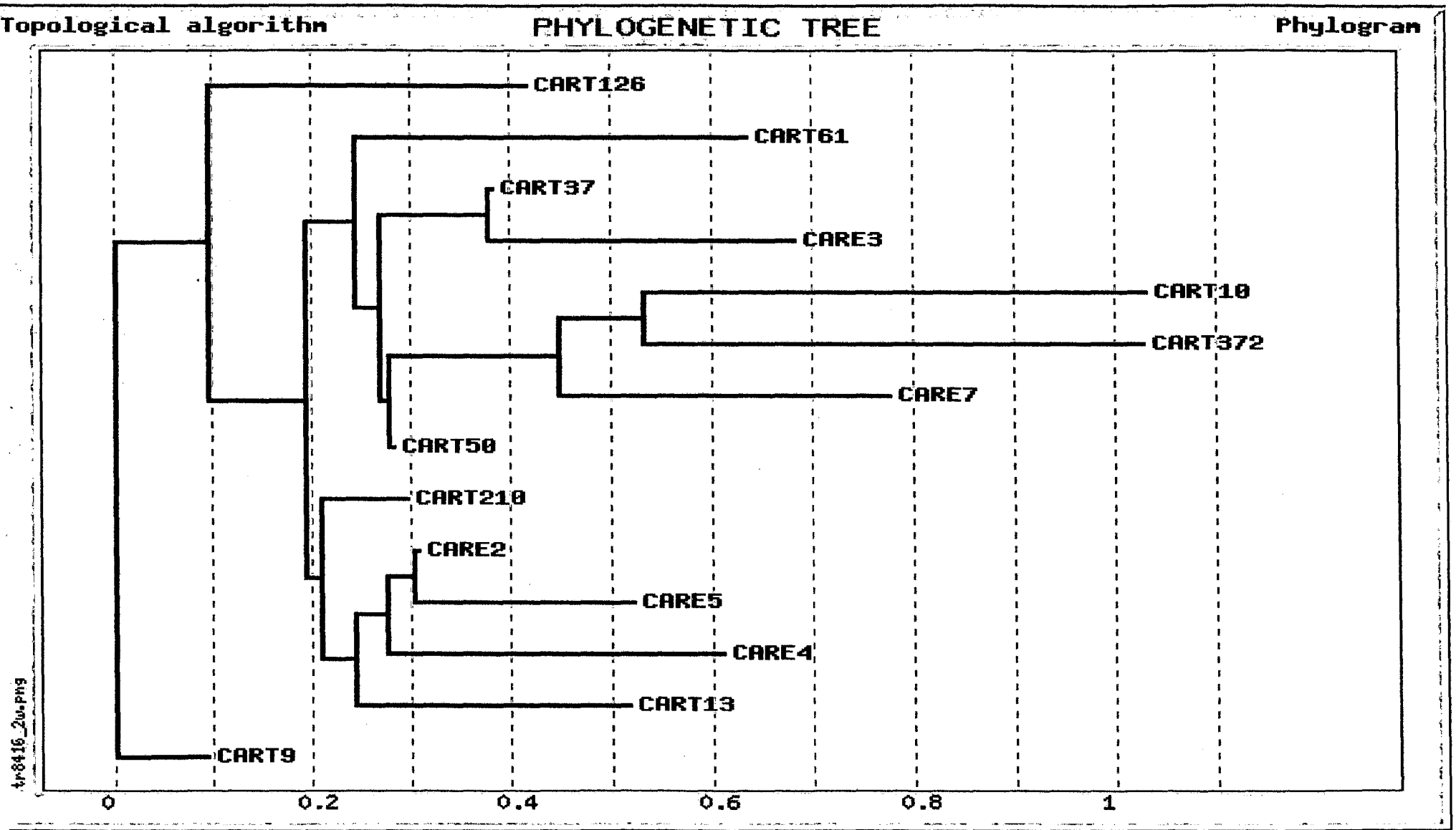


Fig: 17b Phylogenetic tree of predicted peptide sequences of Ty1-copia retrotransposon like sequences of chickpea. Divergences in distance units are indicated by branch lengths. CART- *Cicer arietinum* retrotransposon like sequences, CARE- Transcriptionally active *Cicer arietinum* retroelement like sequences.

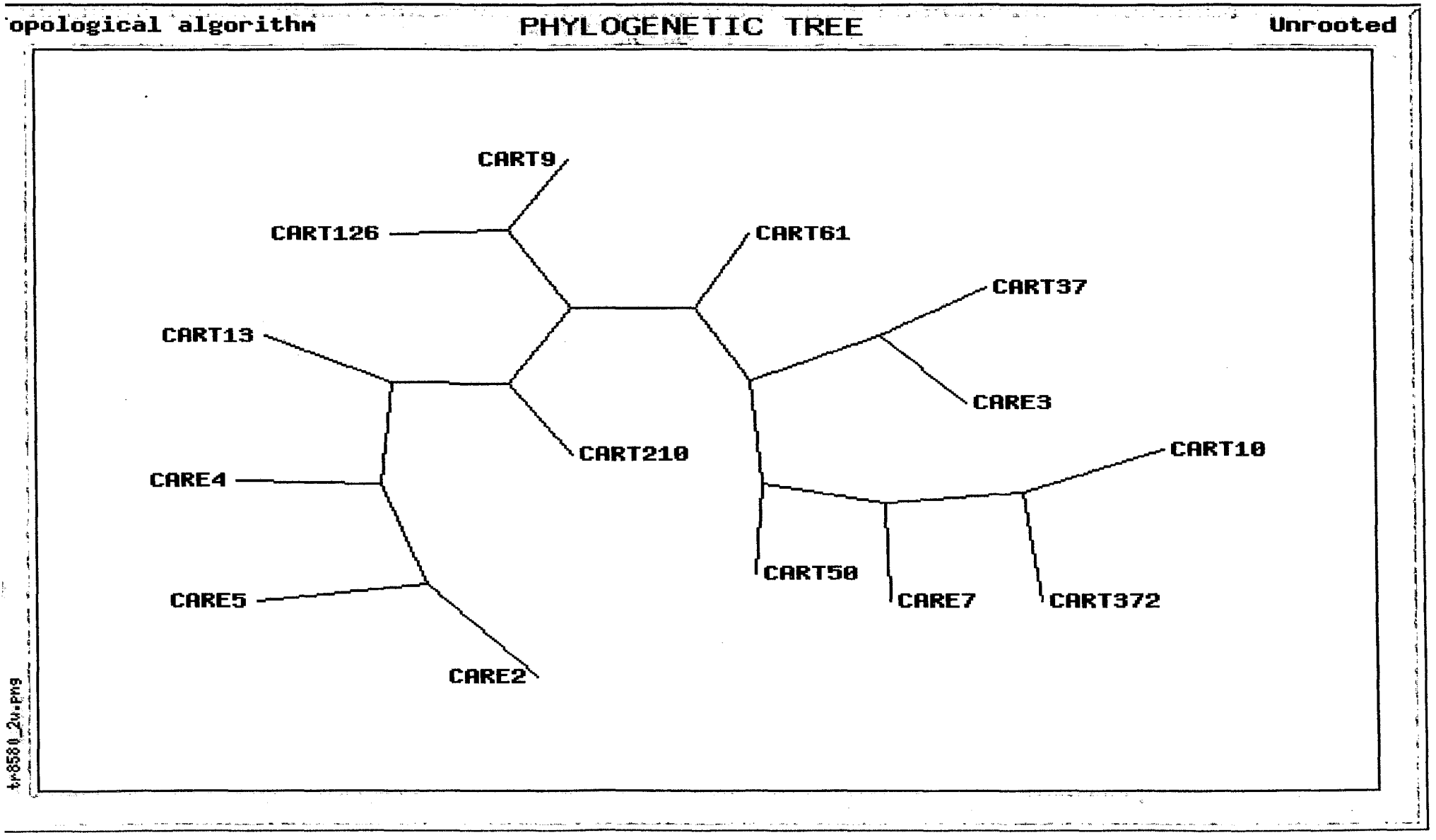


Fig: 17c Phylogenetic tree of predicted peptide sequences of Ty1-copia retrotransposon like sequences of chickpea. Divergences in distance units are indicated by branch lengths. CART- *Cicer arietinum* retrotransposon like sequences, CARE- Transcriptionally active *Cicer arietinum* retroelement like sequences.

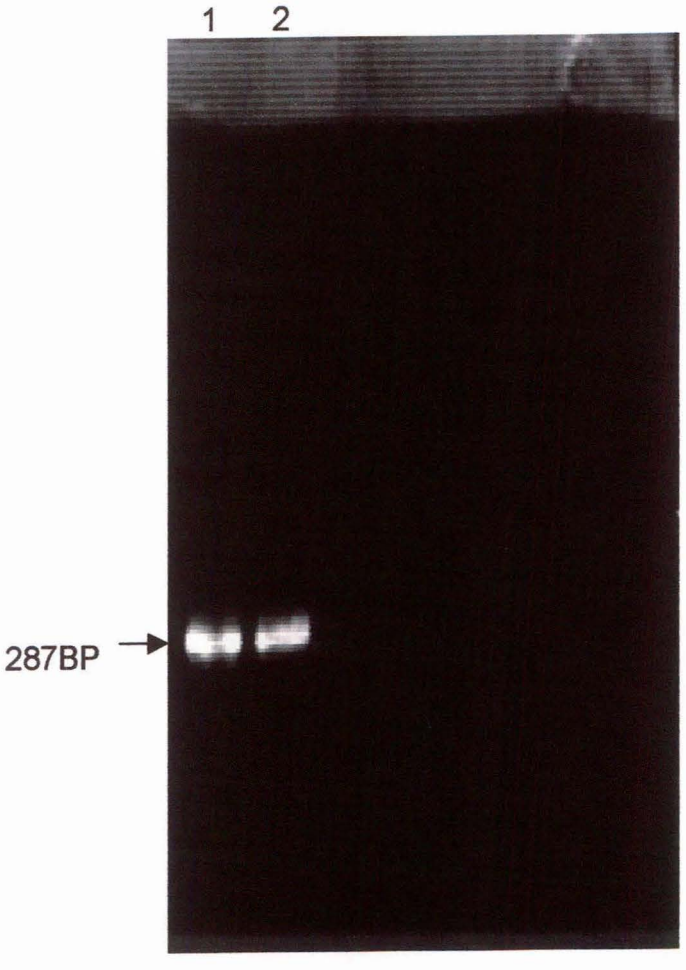


Fig: 18
Amplified products of 10^6- fold dilutions of two different chickpea PCR clones

five RTase mRNAs were found to be transcribed against desiccation stress (a stage between temporary and permanent wilting of plants when kept without water at room temperature) in chickpea (*Cicer arietinum*). These are preliminary observations. The authenticity of these observations is yet to be confirmed following rigorous experimentation.

The comparison of all sub-group representatives with RT sequences of five transcriptionally active *Ty*1-*copia* retrotransposons CARE2, CARE3, CARE4, CARE5 and CARE7 was done to show their phylogenetic relationships and capabilities of being functional (Fig: 17a, b, c). Only sub-group named CART13 showed 84% and 87% amino acid homology to CARE2 and CARE5 respectively, otherwise it was below 70% and in some cases as low as 19%. Thus there was expected little hope that, CART13 sub-group sequences might be related to functional retrotransposons, other sequences would be considered as partial *pol* sequences.

Although utmost care was taken in selection of priming regions but seven sequences (CART5, CART6, CART24, CART81, CART186, CART204 and CART273) with very little non-significant homology towards partial pol sequences and sixteen kinase like partial sequences were found during translated BLASTX analysis (Fig: 12). The occurrence of kinase like sequences might be due non-specific binding of primes or because the primers were degenerate so some nucleotide combination of primers might have affinity towards kinase gene sequences. The kinase like partial sequences did not show any considerable homology among themselves but showed homology to *Arabidopsis* serine kinase (Fig: 16a, b). Thus due to absence of any significant importance these sequences were kept out.

Chapter 5: Discussion

A significant fraction of plant genome is accounted for by presence of mobile genetic elements (*Arabidopsis* Genome Initiative, 2000). These elements on the basis of their transpositional characteristics could be classified into class I and class II transposable elements. The class II (Transposons) elements transpose via DNA intermediate as a "cut and paste" mechanism catalyzed by the element encoded transposes. The class I elements (retroelements or retrotransposons) move via RNA intermediate as a "copy and paste" mechanism (Feschotte *et al.*, 2002 review). The class I elements includes retrotransposons with LTRs (Long Terminal Repeats), non-LTR retrotransposons, LINEs (Long Interspersed Nuclear Elements), SINEs (Short Interspersed Nuclear Elements). LTR-retrotransposons contain LTRs (long terminal repeat) at their termini, a *gag* ORF (Open reading frame) and a polymerase (*pol*) ORF encoding protease, endonuclease, reverse transcriptase and RNase H. PBS (primer finding site) 5′ to gag and PPT (poly purine tract) 3′ to RNase H are also found. The LTR retrotransposons, depending on the basis of the order of the internal domains are classified into *Ty*-copia or *Ty*-gypsy groups. The endonuclease (integrase) domain is positioned 5′ to the reverse transcriptase domain in *copia* group, while in the *gypsy* group it is located 3′ to the reverse transcriptase domain. Now it has been realized that the retroelements constitutes the most abundant and widespread class of transposable elements in plants (Bennetzen *et al.*, 1996, 2002). The presence of retroelements in high copy number in heterogeneous populations, their dispersion through-out the genome and their insertion into new genomic sites without losing the parent copy are some of the properties of retrotransposons, which make them particularly suitable candidate for being utilized as molecular tools in DNA fingerprinting, genetic linkage mapping, phylogenetic studies and molecular breeding. They are also implicated in genome expansion. There exist as many as 2109 retroelements in *Arabidopsis* genome despite its small size (*Arabidopsis* Genome Initiative, 2000). The "C-value paradox" i.e. the lack of correlation between sequence complexity and functional complexity of

genomes can largely be explained by taking the account of the contribution of retroelements to genomes. The success of molecular breeding hinges upon good genetic linkage mapping data and identification of markers closely linked to genes influencing important agronomic traits. Features of retrotransposons like high copy number in highly heterogeneous populations, dispersal throughout the genome, insertion into new genomic sites without losing the parental copies and usually irreversible insertions make them suitable candidates for generating molecular markers in various organisms including plants. 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. The LTRs of retrotransposons are highly conserved regions, which are being exploited for primer designing to develop retrotransposon-based molecular markers. They have been used as DNA markers to study biodiversity in maize, pea and barley and to generate genetic linkage maps in barley, oat and pea. Several techniques such as sequence-specific amplified polymorphism (S-SAP), Inter-retrotransposon amplified polymorphism (IRAP), Retrotransposon-micro-satellite amplified polymorphism (REMAP) and Retrotransposon Based Insertional Polymorphism (RBIP) have emerged during last few years based on LTRs conservancy (Kumar *et al.*, 2001;Kalendar *et al.*, 1999). In legumes AFLP (Amplified Fragment Length Polymorphism) markers can not be proved good for molecular marker assisted breeding programs due to less heterogeneity among their genomes. So the retrotransposons may be used as molecular markers because they are more or less distributed through-out the whole genomes. The genome size of chickpea is 738 Mb and it is expected that retroelements account for largeness of genome. Since retroelements are present in distinct multiple groups, however most of them contain an internal conserve reverse transcriptase domain.

The reverse transcriptase nucleotide sequence is expected to be ~ 300 bp as is cited in various research articles. Also reverse transcriptase sequences commonly have some strongly conserved peptide motifs like LYVDDMDP

(Voytas and Ausubel, 1988). Thus taking this advantage the downstream primer region chosen encodes the LYVDDMDP peptide motif. The upstream primer region chosen encodes the less conserved sequences DVKTAFLHG, and selected in order to isolate or amplify heterogenous population of reverse transcriptases. Only a few members of Ty1-copia group of retrotransposons are reported to be functional. So we focused our efforts in search of Ty1-copia retrotransposon like sequences. By keeping this in mind we selected primers of most strongly conserved regions which was reverse transcriptase region of retrotransposons. The degree of conservation of these primers was considered to be responsible for exclusion of amplification of other retroelements, including gypsy retrotransposon like sequences. The genomic DNA was amplified using these RT primers, not mRNA, because under normal conditions retrotransposons are non-functional. Ty1-copia retrotransposons are expressed only under different types of stress conditions, also all retrotransposons do not express under stress conditions (Beguiristain *et al.*, 2001: Grandbastein, 1998: Hirochika et al., 1993). This is due to the fact that most retrotransposons have become defective by mutations or by insertion of other transposable elements (Bennetzen, 2000 review).

The *gag* region sequences also have conserved sequences but reverse transcriptase (RTase) has advantage over *gag*, being highly conserved then *gag*. So RTase is better choice for PCR amplification than *gag*. Also there exists a tremendous heterogeneity among RTase population, which makes them suitable candidates for isolation and detection of a heterogenous population of retrotransponsons. The heterogeneity among *gag* population is not of much significance.

A large population of the *Ty*l-*copia* group of retrotransposons have been discovered in a diverse collection of eukaryotes, including fungi, animals and plants (Laten, 1999; Linares *et al.*, 2001). In this study we isolated 43 reverse transcriptase sequences (partial *pol* sequences of *Ty*1-copia retrotransposons). We have shown that these sequences are comprised of a very heterogeneous collection of reverse transcriptase sequences. Because the heterogeneity was so

unexpected, that in order to check the reliability of experiment and to eliminate the possibility of artefact, control experiments were performed. Fortunately, all sequences belongs to chickpea partial *pol* regions, as is confirmed by controls. The results interpreted here show that there exists diverse types of *Tyl-copia* group retrotransposon in chickpea than were known to exist in the chickpea prior to this study (Sant *et al.*, 2000). The isolation of 43 different chickpea RTase sequences from 43 randomly picked sub-clones suggests that there exist many more RTase sequences. Furthermore, not all the partial *pol* sequences of Tyl-*copia* group retrotransposons would have been be amplified from the chickpea genome by our choice of PCR primers, because the upstream primer is less conserved, thus there are chances that this upstream primer would not have recognized many of the RTase sequences.

The degeneracy of primers was responsible for amplification of heterogeneous population of reverse transcriptase sequences. The heterogeneity among chickpea retrotransposons exceeds the heterogeneity seen in *D. melanogaster*, which has more *Tyl-copia* group elements per genome than does *Arabidopsis*. So it is assumed that plant *Tyl-copia* group retrotransposons are inherently more susceptible to sequence variation than their counterparts in *Drosophila*. Thus it seems that the special features responsible for great variation in plant genomes are absent in *Drosophila* and *yeast* (Flavell et al., 1994). The most important features in plants are totipotency and the capacity of development of reproductive or somatic tissue from meristematic tissue throughout the life cycle. In this way the mutations occurring in meristematic cells might have passed to germ cell because plants are tolerant to individual cell damage. In case of animals mosaicism of germline gets established early during embryogenesis and the germ cells may be better protected against mutations. Also another feature of plants may be their ploidy and supernumerary chromosomes. These features make plant genomes more tolerant to chromosome alterations, by the principle of compensation.

A very good amino acid homology in range of 46-76% of chickpea RT sequences towards *Ty*1-copia RTase sequences of different plants confirmed

that all 43 sequences isolated and characterized by us are reverse transcriptase sequences of Ty1-copia retrotransposons. Since all sequences are different so these belongs to different Ty1-copia retrotransposons. Thus we can consider that there exists a large heterogeneous population of Ty1-copia retrotransposons in chickpea. The control PCR reactions (with 10^6 fold dilutions of two different chickpea PCR clones) had been performed to prove the reliability of our results. The retrotransposon sequences of chickpea are in some way different from other systems except plants, because during BLASTX analysis there was hardly any homologous sequence other than plant origin, if it was there it had less than 10% amino acid homology towards chickpea RT sequences.

The presence of 4-21 stop codons in all six reading frames of most of these sequences supports their defective and non-functional nature. Other sequences also which do not bear stop codon in at least one reading frame, do not show any significant homology towards transcriptionally active retrotransposons except CART77. The occurrence of stop codons in one third of total number of retrotransposon sequences should be considered for their translational non-functionality also. The presence of more than one translation stop codon in every reading frame may be as a result of mutations and epigenetic movements. It might be possible that all retrotransposons have common origin and during the course of evolution due to insertion of retroelements in their own copy or relative sequence have given rise to a great heterogeneity among retrotransposons. Other two third sequences do not have stop codon at least in one of their reading frame. But RTase sequences without stop codon may not be considered as the only factor responsible for the functionality of a particular retrotransposon. Reverse transcriptase is one of the important agent, which is involved in reverse transcription of RNA into complementary DNA. It is also observed that the copy number of retrotransposons increases many folds upon stress induction thus the retrotransponsons seems to play some defensive role against various stresses. Also due this reason retrotransposons have played key role during evolution by

restructuring genome by their property of retrotransposition. One crude reason for the non-functionality of these elements might be their no advantages in absence of stress.

Translated sequences were compared by keeping in mind the degeneracy of nucleotides, so that the actual encoded peptides could be analyzed. All RT sequences were derived from common parent sequences, but they diversified during course of evolution. Their heterogeneity might have developed in response to various biotic and abiotic stresses. The 80-96% intra sub-group amino acid homology strongly supported their recent diversification but 45-80% inter sub-group amino acid homology support their long inter sub-group diversification.

The degree of amino acid divergence criteria for classification of RT sequences was very well supported by their phylogenetic trees. All sequences of a sub-group were shown to originate from common main branch. Rooted, un-rooted and slanted different types of phylogenetic trees also agree with their common origin. Although all sub-groups may have common origin but they had diversified up to such an extent that they resembled to other plant sequences as much as to chickpea RT sequences. So these sequences fall into different sub-groups.

The degree of amino acid divergence among these sub-groups may be considered as the criteria to know their phylogenetic relationships i.e. how much far they have come independently. Sequences evolving together showed least degree of amino acid divergence. The existence of unexpected large amino acid divergence between CART372 and CART381 could be due to their independent evolution. Although these two sequences were considered to be originated from same sequence source but they might have diversified and evolved in response to different types of environmental stress conditions.

Apart from retrotransposon heterogeneity, another interesting observation in our work is that clone CART 77 (Acc. No. AJ 535867) possesses more than 85% homology to *copia*-like RTase region of *Tto*1 element of tobacco in its first reading frame without any stop codon (See

alignment of CART 77 with *Tto* 1). The translated CART77 sequence does not bear any stop codon in its first 5′-3′ reading frame, the same is true with active retrotransposon *Tto*1 RTase of tobacco (Hirochika et al., 1996). Eighty five percent homology of CART77 with *Tto*1 supports its functional nature strongly. Out of 89 amino acids of CART77, 70 are identical, 7 conserved and 2 semi-conserved with Tto1 translated RTase mRNA. This observation strongly recommends the translational functionality of CART 77 reverse transcriptase of Chickpea (*Cicer arietinum*).

The comparison of all chickpea CART series of sequences with stress induced transcriptionally active RTases showed less than 70% amino acid homology except CART13 sub-group, which showed 84-87% homology. Thus members of CART13 might be transcriptionally active. But all transcriptionally active retrotransposons may not be translationally functional, because most commonly these sequences are interrupted by a number of stop codons.

The occurrence of non-specific kinase like sequences may be explained by considering non-specific binding of primers at the genomic DNA.

Retrotransposons are class of dispersed middle repetitive sequences which have contributed to the genetic diversity of their host species. Transposition and homologous recombination between retrotransposons have been reported to be involved in generating variability among genomes. The diversity generated by the elements has been studied in barely, maize and pea (Ellis et al., 1998; Kalendar et al., 1999; Purugamnan and Wessler, 1995). Due to this feature, retrotransposons have been used as efficient DNA fingerprinting probes in some plant species and have further been exploited as DNA markers to generate genetic linkage maps in barley and pea (Ellis et al., 1998; Kumar et al., 1997; Wang et al., 1999).

Retrotransposons have been reported to be capable of generating large populations in a relatively short evolutionary course of time as they have a replicative mode of transposition. Plant genomes can accumulate large amounts of DNA, so it is possible that during evolution retrotransposons have proliferated into larger populations from a few active elements in the host

genome (Kumar and Bennetzen, 1999). Most Tyl-*copia* group retrotransposon are transcriptionally inactive, thus thought to be fixed in their genomic locations. Active Tyl-*copia* elements have been found in tobacco (*Tnt* 1, *Tto* 1), barely, wheat, Oat (*OARE* 1) and rye (*BARE* 1) (Flavell et al., 1997). *In situ* hybridization studies have shown that *Tnt* 1 and *BARE* 1 elements are located in the euchromatic region, while in *A. Thaliana* and *C. arietinum,* the *Tyl-copia* elements are clustered in the genetically inactive centromeric hetrochromatic region (Brandes et al., 1997). However, it has also been reported that the transcriptional activity of such elements could be induced under several biotic and abiotic stresses (Grandbastein, 1998).

Since each retrotransposon has a unique phylogenetic history, these elements have been used to sort-out phylogenetic relationships in cereal plants. The phylogenetic relationship among wheat, rice, and maize have been studied on the basis of Tyl-*copia* RTase domain (Matsuoka and Tsunewaki, 1999). Similarly our data could be used with RTase sequences from other legumes to reveal their evolutionary relationships.

Although the priming regions selected were strongly conserved, but seven sequences (CART5, CART6, CART24, CART81, CART186, CART204 and CART273) with very little non-significant homology towards partial pol sequences and sixteen kinase like partial sequences were found during translated BLASTX analysis (Fig: 12). Also the degree of conservancy of upstream primer encoding DVKTAFLHG peptide motif was less in comparison to down stream primer (Voytas and Ausubel, 1988). The occurrence of kinase like sequences might be due non-specific binding of primes or because the primers were degenerate so some nucleotide combination of primers might have affinity towards kinase gene sequences. The kinase like partial sequences did not show any considerable homology among themselves but showed some homology to *Arabidopsis* serine kinase (Fig: 16a, b). Thus due to absence of any significant importance these sequences were kept out.

The transcriptionally active RTase sequences CARE2, 3, 4, 5 and CARE7 have been isolated. These sequences are found to be transcribed

against desiccation stress in chickpea (*Cicer arietinum*). These sequences when translated using computer software are found to have few translational stop codons. The sequences CAER2 and CARE5 do not bear any stop codon in their 2^{nd} 5′-3′ reading frames. Thus the retrotransposons to which these sequences belong are expected to be transpositionally functional. The work is in progress to isolate their complete retrotransposon. The observations are yet to be proved experimentally.

Chapter 6: Summary and conclusions

The knowledge of mobile genetic elements (Transposable elements) began with the discovery of transposons by Barbara McClintok. These rapidly accumulating transposons are responsible for dynamicity of the complex genomes. A significant fraction of plant genome is accounted for by presence of mobile genetic elements. Transposable elements on the basis of their transpositional characteristics have been classified into class I and II. The class II elements known as DNA transposable elements transpose via DNA intermediates a "cut and paste" mechanism catalyzed by the element encoded transposase. They constitute autonomous and non-autonomous transposons and a family of small elements (100-500 bps) called MITEs (Miniature Inverted repeat Transposable Elements). The class-I elements include LTR-retrotransposons, non-LTR-retrotransposons, long interspersed nuclear elements (LINEs) and short interspersed nuclear elements (SINEs). These elements or retrotransposons transpose via an RNA intermediate by a "copy and paste mechanism". LTR-retrotransposons contain LTRs (long terminal repeat) at their termini, a *gag* ORF (Open reading frame) and a polymerase (*pol*) ORF encoding protease, endonuclease, reverse transcriptase and RNase H. PBS (primer finding site) 5′ to gag and PPT (poly purine tract) 3′ to RNase H are also found. The LTR-retrotransposons have been divided into *Ty*1-*copia* and *Ty*3-*gypsy* groups on the basis of sequence homology and the order of the internal domains. The endonuclease domain is positioned 5′ to the reverse transcriptase domain in *copia* group, while in the *gypsy* group it is located 3′ to the reverse transcriptase domain.

There exist as many as 2109 retroelements in *Arabidopsis* genome despite its small size. The "C-value paradox" i.e. the lack of correlation between sequence complexity and functional complexity of genomes can largely be explained by taking the account of the contribution of retroelements to genomes.

The success of molecular breeding hinges upon good genetic linkage mapping data and identification of markers closely linked to genes influencing

important agronomic traits. Features of retrotransposons like high copy number in highly heterogeneous populations, dispersal throughout the genome, insertion into new genomic sites without losing the parental copies and usually irreversible insertions make them suitable candidates for generating molecular markers in various organisms including plants. 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.

In legumes AFLP (Amplified Fragment Length Polymorphism) markers can not be proved good for molecular marker assisted breeding programs due to less heterogeneity among their genomes. So the retrotransposons may be used as molecular markers because they are more or less distributed throughout the whole genomes.

The genome size of chickpea is 738 Mb and it is expected that retroelements account for largeness of the genome. Since retroelements are present in distinct multiple groups, however most of them contain a internal conserve reverse transcriptase domain. The reverse transcriptase nucleotide sequence is expected to be ~ 300 bp as is cited in various research articles. Also reverse transcriptase sequences commonly have some strongly conserved peptide motifs like LYVDDMDP. The objective of research was to isolate and characrerize retroelements from the chickpea (*Cicer arietinum*) genome. Thus taking the advantage of conserved regions primers were designed. The downstream primer region chosen encodes the LYVDDMDP peptide motif. The upstream primer region chosen encodes the less conserved sequences DVKTAFLHG, and selected in order to isolate or amplify heterogenous population of reverse transcriptases.

Altogether 43 reverse transcriptase (partial pol) sequences were found from restriction digestion analysis of total 400 transformed white bacterial colonies (see list of accession numbers). Although all 43 sequences show significant homology to the reverse transcriptase region of *Tyl-copia* of other plants but none of two sequences are identical. All 43 sequences were translated using computer program and then compared among themselves and

with the partial *pol* sequences of *Tyl-copia* retrotransposons from other plants. Analysis of their alignment and phylogenetic tree drawn show the existence of extreme heterogeneity in *Tyl-copia* retrotransposons of chickpea. Based on their phylogenetic relationships all 43 sequences are categorized in 9 sub-groups. The degree of divergence observed is around 20-55% between the members of different sub-groups but it is 4-20% between the members of the same sub-group.

One third of these reverse transcriptase sequences are considered to belong to the defective *Tyl-copia* retrotransposons of chickpea. The presence of stop codons could be considered responsible only for their hindrance to translation. So the transcription from these retrotransposons could be possible. The clone CART 77 (Acc. No. AJ 535867) is found to have 85.36% homology to *Tto*1 RTase, which belongs to functional retrotransposon of *N. tabaccum*. The comparison of CART 77 to *Tto*1 supports the possibility that CART 77 might be functional.

Chapter 7: References

Arabidopsis Genome Initiative (2000) Nature (London). **408**: 796-815.

Aravin AA (2001) Double-stranded RNA mediated silencing of genomic tandem repeats and transposable elements in the Drosophila Melanogaster germline. Curr.Biol. **11**:1017-1027.

Avramova Z, Sanmiguel P, Georgieva E, Bennetzen JL (1995) Matrix attachment regions and transcribed sequences within a long chromosomal continuum containing maize Adh1. Plant Cell. 7:1667-80.

Bachmair A, Luschnig C, Hess M, Pusch O, Brookman J (1995) The gag homologue of retrotransposon Ty1 assembles into spherical particles in Escherichia coli. Eur J Biochem . **228**:739-44.

Baltimore D (1970) RNA-dependent DNA polymerase in virions of RNA tumor viruses. Nature. **226**:1209-1211.

Beguiristain T, Grandbastien MA, Puigdomenech P and Casacuberta JM (2001) Three *Tnt1* subfamilies show different stress-associated patterns of expression in tobacco. Consequence of retrotransposon control and evolution in plants. Plant Physiol. **127**: 212–221.

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

Bennetzen JL, SanMiguel P, Gaut BS, Tikhonov A, Nakajima Y(1998)The paleontology of intergene retrotransposons of maize. Nat Genet. **20**:43-5.

Bennetzen JL (1996) The contribution of retroelements to plant genome organization, function and evolution. Trends Microbiol. **4**: 347-353.

Brandes JL, Harrison JS, kumar A, Kamm A, Kubis S, Doudrick RL, Schimdt T (1997) Comparative analysis of the chromosomal and genomic organization of Ty1-copia like retrotransposons in pteridophytes, gymnosperms and angiosperms. Plant Mol Biol. **33**:11-21.

Bureau TE and Wessler SR (1994) Stowaway: a new family of inverted repeat elements associated with the genes of both monocotyledonous and dicotyledonous plants. Plant Cell. **6**:907-16.

Cresse AD, Hulbert SH, Brown WE, Lucas JR, Bennetzen JL (1995) Mu1-related transposable elements of maize preferentially insert into low copy number DNA. Genetics. **140**:315-24.

Curico MJ and Garfinkel DJ (1999) New lines of host defense: inhibition of Ty1 retrotransposition by Fus3p and NER/TFIIH. Trends Genet. **15**:43-5. Review.

Djikeng A, Shi H, Tschudi C, Ullu E (2001) RNA interference in Trypanosoma brucei: cloning of small interfering RNAs provides evidence for retroposon-derived 24-26-nucleotide RNAs. RNA. **7**:1522-30.

Doolittle RF, Feng DF (1990) Nearest neighbour procedure for relating progressively aligned amino acid sequences. Meth Enzymol. **183**:659-669.

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

Ellis TH, Poyser SJ, Knox MR, Vershinin AV and Ambrose MJ (1998) Polymorphism of insertion sites of *Ty1*-copia class retrotransposons and its use for linkage and diversity analysis in pea. Mol. Gen. Genet. **260**: 9–19.

Emori Y, Shiba T, Kanaya S, Inouye S, Yuki S and Saigo K (1985) Determination of the nucleotide sequences of copia and copia related RNA in Drosophila VLP. Nature. **315**:773-776.

Feng DF, Doolittle RF (1990) Progressive alignment and phylogenetic tree construction of protein sequences. Meth in Enzymol. **183**:375-387.

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**:329-41. Review.

Feuerbach F, Lucas H (2001) The protease and reverse transcriptase of the tobacco LTR retrotransposon Tnt1 are enzymatically active when expressed in Escherichia coli. Plant Mol Biol. **46**:481-9.

Flavell AJ, Knox MR, Pearce SR and Ellis TH (1998) Retrotransposon-based insertion polymorphisms (RBIP) for high throughput marker analysis. Plant J. **16**: 643–650.

Flavell AJ, Pearce SR, Heslop-Harrison JS, Kumar A (1997) The evolution of Ty1-copia group retrotransposons in eukaryote genomes. Genetica. **100:** 185-195.

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

Flavell AJ, Smith DB, Kumar A (1992a) Extreme hetergeneity of Ty1-copia group retrotransposons in plants. Mol Gen Genet. **231**:233-242.

Flavell AJ (1992b) Ty1-copia group retrotransposons and the evolution of retroelements in the eukaryotes. Genetica. **86**:203-14.

Flavell AJ and Smith DB (1992c) A Ty1-copia group retrotransposon sequence in a vertebrate. Mol. and Gen Gent. **233:**322-326.

Gaucher EA, Majumdar A, Laten HM (1998) SIRE-1, a copia/Ty1-like retroelement from soybean, encodes a retroviral envelope-like protein.Proc Natl Acad Sci U S A. **95**:6897-902.

Gaut BS, Le Thierry d'Enneguin M, Peek AS, Sawkins MC (2000) Maize as a model for the evolution of plant nuclear genomes. Proc Natl Acad Sci U S A. **97** :7008-15. Review.

Grandbastien MA, Beguiristain T, Puigdomenech P, Casacuberta JM (2001) Three Tnt1 subfamilies show different stress-associated patterns of expression in tobacco. Consequences for retrotransposon control and evolution in plants.Plant Physiol. **127**:212-21.

Grandbastien, MA (1998) Activation of plant retrotransposons under stress conditions. Trends Plant Sci. **3**:181-187.

Grandbastien MA, Lucas H, Morel JB, Mhiri C, Vernhettes S, Casacuberta JM (1997).The expression of the tobacco Tnt1 retrotransposon is linked to plant defense responses. Genetica. **100** :241-52. Review.

Grandbastien MA (1992) Retroelements in higher plants.Trends Genet. **8**:103-8. Review.

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.

Graves JA, O'Neill MJ, O'Neills RJ (1998) Undermethylation associated with retroelement activation and chromosome remodelling in an interspecific mammalian hybrid. Nature. **393**:68-72.

Gribbon BM, Pearce SR, Kalendar R, Schulman AH, Paulin L, Jack P, Kumar A and Flavell AJ (1999) Phylogeny and transpositional activity of *Ty1*-copia group retrotransposons in cereal genomes. Mol. Gen. Genet. **261**: 883–891.

Heslop-Harrison JS, Brandes A, Taketa S, Schmidt T, Vershinin AV, Kubis S, Kumar A, Pearce SR, Flavell AJ, Harrison GE (1997) The chromosomal distributions of Ty1-copia group retrotransposable elements in higher plants and their implications for genome evolution. Genetica. **100**:197-204.

Higgins DG, Sharp PM (1988) CLUSTAL: A package for performing multiple sequence alignments on microcomputer. Gene **73**:237-244.

Hirochika H, Okamoto H, Kakutani T (2000) Silencing of retrotransposons in arabidopsis and reactivation by the ddm1 mutation. Plant Cell. **12**:357-69.

Hirochika H, Otsuki H, Yoshikawa M, Otsuki Y, Sugimoto K, Takeda S (1996) Autonomous transposition of the tobacco retrotransposon Ttol in rice. Plant Cell. **8**:725-34.

Hirochika H (1993a) Activation of tobacco retrotransposons during tissue culture.EMBO J. **12**:2521-8.

Hirochika H, Hirochika R (1993b) Ty1-copia group retrotransposons as ubiquitous components of plant genomes. Jpn J Genet. **68**:35-46.

Hu W, Das OP, Messing J (1995) Zeon-1, a member of a new maize retrotransposon family. Mol Gen Genet. **248**:471-80.

Jin YK, Bennetzen JL (1989) Structure and coding properties of Bs1, a maize retrovirus-like transposon. Proc Natl Acad Sci U S A. **86**:6235-9.

Johns MA, Mottinger J, Freeling M (1985) A low copy number, copia-like transposon in maize. EMBO J. **4**:1093-1102.

Kalendar R, Tanskanen J, Immonen S, Nevo E and Schulman AH (2000) Genome evolution of wild barley (*Hordeum spontaneum*) by BARE-1 retrotransposon dynamics in response to sharp microclimatic divergence. Proc. Natl. Acad. Sci. USA. **97:** 6603–6607.

Kalendar R, Grob T, Regina M, Suoniemi A, Schulman AH (1999) IRAP and REMAP: two new retrotransposon-based DNA fingerprinting techniques.Theor. Appl. Genet. **33:** 479-532.

Kamm A, Doudrick RL, Heslop-Harrison JS, Schmidt T (1996) The genomic and physical organization of Ty1-copia-like sequences as a component of large genomes in Pinus elliottii var. elliottii and other gymnosperms.Proc Natl Acad Sci U S A. **93**:2708-13.

Kapitonov VV and Jurka J (2001) Rolling-circle transposons in eukaryotes. Proc Natl Acad Sci U S A. **98**:8714-9.

Katsiotis A, Schmidt T, Heslop-Harrison JS (1996) Chromosomal and genomic organization of Ty1-copia-like retrotransposon sequences in the genus Avena.Genome; **39**:410-7.

Kempken F, Kuck U (1998) Transposons in filamentous fungi--facts and perspectives. Bioessays. **8**:652-9. Review.

Kidwell MG, Lisch DR (1997) Hybrid genetics. Transposons unbound.Nature. **393**:22-3.

Kikuchi F, Fukuchi A and Hirochika H (1992) Retrotransposon families in rice. Mol Gen Genet. **233**:209-16.

Kim A, Terzian C, Santamaria P, Pelisson A, Purdhomme N, Bucheton A (1994) Retroviruses in invertebrates: the gypsy retrotransposon is apparently an infectious retrovirus of Drosophila melanogaster.Proc Natl Acad Sci U S A. **91**:1285-9.

Kimura Y, Tosa Y, Shimada S, Sogo R, Kusaba M, Sunaga T, Betsuyaku S, Eto Y, Nakayashiki H, Mayama S (2001) OARE-1, a Ty1-copia retrotransposon in oat activated by abiotic and biotic stresses. Plant Cell Physiol. **42**:1345-54.

Kubis SE, Heslop-Harrison JS, Desel C, Schmidt T (1998a) The genomic organization of non-LTR retrotransposons (LINEs) from three Beta species and five other angiosperms.Plant Mol Biol. **36**:821-31.

Kubis SE, Brandes A, Taketa S, Schmidt T, Kamm A, Schwarzachr T, Vershinin AV, Alkhimova EG, Doudrick RL, Katsiotis A, Kumar A, Pearce SR, Flavell AJ, Harrison GE, Haeslop-Harrison JS (1998b) The chromosomal

distributions of Ty1-copia group retrotransposable elements in higher plants and their implications for genome evolution.Genetica. **100**:197-204.

Knoop V, Unseld M, Marienfeld J, Brandt P, Sunkel S, Ullrich H, Brennicke A (1996) copia-, gypsy- and LINE-like retrotransposon fragments in the mitochondrial genome of Arabidopsis thaliana.Genetics. **142**:579-85.

Kumar A, Hirochika H (2001) Applications of retrotransposons as genetic tools in plant biology. Trends Plant Sci. **6**:127-34. Review.

Kumar A, Bennetzen JL (2000) Retrotransposons: central players in the structure, evolution and function of plant genomes. Trends Plant Sci. **5**:509-510.

Kumar A and Bennetzen JL (1999) Plant retrotransposons. Annu. Rev. Genet. **33:** 479–532.

Kumar A, Pearce SR, McLean K, Harrison G, Heslop-Harrison JS, Waugh R, Flavell AJ (1997) The Ty1-copia group of retrotransposons in plants: genomic organisation, evolution, and use as molecular markers. Genetica. **100**:205-17.

Kumar A, Pearce SR, Pich U, Harrison G, Flavell AJ, Heslop-Harrison JS, Schubert I (1996) The Ty1-copia group retrotransposons of Allium cepa are distributed throughout the chromosomes but are enriched in the terminal heterochromatin.Chromosome Res. **4**:357-64.

Kumekawa N, Ohmido N, Fukui K, Ohtsubo E, Ohtsubo H (2001) A new gypsy-type retrotransposon, RIRE7: preferential insertion into the tandem repeat sequence TrsD in pericentromeric heterochromatin regions of rice chromosomes. Mol Genet Genomics.**265**:480-8.

Kumekawa N, Ohtsubo E, Ohtsubo H (1999a) Identification and phylogenetic analysis of gypsy-type retrotransposons in the plant kingdom. Genes Genet Syst. **74**:299-307.

Kunze R, Becker HA (1996) Binding sites for maize nuclear proteins in the subterminal regions of the transposable element Activator.Mol Gen Genet. **251**:428-35.

Kuwahara A, Kato A, Komeda Y (2000) Isolation and characterization of copia-type retrotransposons in Arabidopsis thaliana. Gene. **244**:127-36.

Lall IP, Maneesha, Upadhyaya KC (2002) Panzee, a copia-like retrotransposon from the grain legume, pigeonpea (Cajanus cajan L.).Mol Genet Genomics. **267**:271-80

Laten HM (1999) Phylogenetic evidence for Ty1-copia-like endogenous retroviruses in plant genomes. Genetica. **107**:87-93.

Linares C, Loarce Y, Serna A, Fominaya A (2001) Isolation and characterization of two novel retrotransposons of the Ty1-copia group in oat genomes. Chromosoma. **110**:115-23.

Lindroth AM, Cao X, Jackson JP, Zilberman D, McCallum CM, Henikoff S, Jacobsen SE (2001) Requirement of CHROMOMETHYLASE3 for maintenance of CpXpG methylation.Science. **292**:2077-80.

Liu B, Wendel JF (2000) Retrotransposon activation followed by rapid repression in introgressed rice plants. Genome. **43**:874-80.

Li ZY, Chen SY, Zheng XW, Zhu LH (2000) Identification and chromosomal localization of a transcriptionally active retrotransposon of Ty3-gypsy type in rice. Genome. **43**:404-8.

Luan DD, Korman MH, Jakubczak JL, Eickbush TH (1993) Reverse transcription of R2Bm RNA is primed by a nick at the chromosomal target site: a mechanism for non-LTR retrotransposition.Cell. **72**:595-605.

Lucas H, Feuerbach F, Kunert K, Grandbastein MA, Caboche M (1995) RNA-mediated transposition of the tobacco retrotransposon Tnt1 in Arabidopsis thaliana.EMBO J. **14**:2364-73.

Luschnig C, Bachmair A (1997) RNA packaging of yeast retrotransposon Ty1 in the heterologous host, Escherichia coli. Biol Chem. **378**:39-46.

Manninen I, Schulman AH (1993) BARE-1, a copia-like retroelement in barley (Hordeum vulgare L.).Plant Mol Biol; **22**:829-46.

Manninen O, Kalendar R, Robinson J,Schulman AH (2000) Application of BARE-1 retrotransposon markers to map a major resistance gene for net blotch in barley. Mol. Gen. Genet. **264**:325-334.

Matsuoka Y, Tsunewaki K (1999) Evolutionary dynamics of Ty1-copia group retrotransposons in grass shown by reverse transcriptase domain analysis. Mol Biol Evol. **16**:208-17.

McClintock B (1967) Components of action of the regulators *Spm* and *Ac*. Carnegie Inst. Wash. Yearb. **64:** 527-534.

McClintock B (1951) Chromosome organization and genic expression. Cold Spring Harbor Symp. Quant. Biol. **16:** 13–47.

McClure MA (1991) Evolution of retroposons by acquisition or deletion of retrovirus-like genes.Mol Biol Evol. **8**:835-56.

Melayah D, Bonnivard E, Chalhoub B, Audeon C, Grandbastien MA (2001) The mobility of the tobacco Tnt1 retrotransposon correlates with its transcriptional activation by fungal factors. Plant J. **28**:159-68.

Messing J, Hu W, Das OP (1995) Zeon-1, a member of a new maize retrotransposon family.Mol Gen Genet. **248**:471-80.

Meyers BC, Tingey SVand Morgante M.(2001) Abundance, distribution and transcriptional activity of repetitive elements in maize genome.Genome Res. **11**: 1660-1676.

Meyers BC, Chin DB, Shen KA, Sivaramakrishan S, Lavelle DO, Zhang Z, Miichelmore RW (2000) The major resistance gene cluster in lettuce is highly duplicated and spans several megabases.Plant Cell. **10**:1817-32.

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.

Miura (2001) Mobilization of transposons by mutation abolishing full DNA methylation in Arabidopsis. Nature. **411**: 212- 214.

Moffat AS (2000) Transposons help sculpt a dynamic genome. Science; **289:** 1455–1457.

Moore G, Lucas H, Batty N, Flavell R (1991) A family of retrotransposons and associated genomic variation in wheat.Genomics. **10**:461-8.

Moran JV, Lutz-Prigge S, Gilbert N (2002) Genomic deletions created upon LINE-1 retrotransposition.Cell. **110**:315-25.

Moran JV (1999) Human L1 retrotransposition: insights and peculiarities learned from a cultured cell retrotransposition assay.Genetica. **107**:39-51. Review.

Morgante M, Tinget SV, Meyers BC (2001) Abundance, distribution, and transcriptional activity of repetitive elements in the maize genome.Genome Res. **11**:1660-76.

Mount SM, Rubin GM (1985) Complete nucleotide sequence of the Drosophila transposable element copia: homology between copia and retroviral proteins. Mol Cell Biol. **5**:1630-8.

Nakayashiki H, Matsuo H, Chuma I, Ikeda K, Betsuyaku S, Kusaba M, Tosa Y, Mayama S (2001) Pyret, a Ty3/Gypsy retrotransposon in Magnaporthe grisea contains an extra domain between the nucleocapsid and protease domains.Nucleic Acids Res. **29**:4106-13.

Ohtsubo H, Kumekawa N, Ohtsubo E (1999b) RIRE2, a novel gypsy-type retrotransposon from rice. Genes Genet Syst. **74**:83-91.

Ohtsubo H, Umeda M, Ohtsubo E (1991) Organization of DNA sequences highly repeated in tandem in rice genomes. Jpn J Genet. **66**:241-54.

Palmgren MG (1994) Capturing of host DNA by a plant retroelement: Bs1 encodes plasma membrane H(+)-ATPase domains. Plant Mol Biol. **25**:137-40.

Pardue ML, Danilevskaya ON, Traverse KL, Lowenhaupt K (1997) Evolutionary links between telomeres and transposable elements. Genetica. **100**:73-84. Review

Pearce SR, Knox M, Ellis TH, Flavell AJ and Kumar A (2000) Pea *Ty1*-copia group retrotransposons: transpositional activity and use as markers to study genetic diversity in *Pisum*. Mol. Gen. Genet. **263:** 898–907.

Pearce SR, Stuart-Rogers C, Knox MR, Kumar A, Ellis TH and Flavell AJ (1999) Rapid isolation of plant *Ty1*-copia group retrotransposon LTR sequences for molecular marker studies. Plant J. **19:** 711–717.

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 chromosomsl location. Mol. Gen. Genet. **250:** 305-315.

Pelissier T, Tutois S, Deragon JM, Tourmente S, Genestier S, Picard G (1995) Athila, a new retroelement from Arabidopsis thaliana. Plant Mol Biol. **29**:441-52.

Peterson-Burch BD, Wright DA, Laten HM, Voytas DF (2000)Retroviruses in plants?Trends Genet. **16**:151-2. Review.

Pouteau S, Spielmann A, Meyer C, Grandbastien MA, Caboche M (1991) Effects of Tnt1 tobacco retrotransposon insertion on target gene transcription. Mol Gen Genet. **228**:233-9.

Pouteau S, Audeon C, Casacuberta JM, Grappin P, Lucas H, Moreau C, Grandbastien MA (1994) Functional analysis of the tobacco Tnt1 retrotransposon. Genetica. **93**:181-9. Review.

Purugganan MD, Wessler SR (1995) Transposon signatures: species-specific molecular markers that utilize a class of multiple-copy nuclear DNA. Mol Ecol. **4**:265-9.

Razin A, Riggs AD (1980) DNA methylation and gene function. Science. **210**:604-10. Review.

Richards EJ, Jeddeloh JA, Stokes TL (1999) Maintenance of genomic methylation requires a SWI2/SNF2-like protein. Nat Genet. **22**:94-7.

Rogers SA, Pauls KP (2000) Ty1-copia-like retrotransposons of tomato (Lycopersicon esculentum Mill.). Genome. **43**:887-94.

Sabl JF, Henikoff S (1996) Copy number and orientation determine the susceptibility of a gene to silencing by nearby heterochromatin in Drosophila. Genetics. **142**:447-58.

Sambrook, J., Fritsh, F., Maniatis, T (1989) Molecular Cloning: A Laboratory Manual. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

Sanger F, Nicklen S, Coulson AR (1977) DNA sequencing with chain-terminating inhibitors. PNAS USA. **74:**5463-5467

SanMiguel P, Tikhonov A, Jin YK, Motchoulskaia N, Zakharov D, Melake-Berham A, Springer PS, Edwards KJ, Lee M, Avramova Z, Bennetzen JL (1996) Nested retrotransposons in the intergenic regions of maize genome. Science. **274:**765-768.

SanMiguel P, Gaut BS, Tikhonov A, Nakajima Y and 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**:157-66.

Schmidt T (1999) LINEs, SINEs and repetitive DNA: non-LTR retrotransposons in plant genomes. Plant Mol Biol. **40**:903-10. Review.

Schulman AH, Nevo E, Kalendar R, Tanskanen J, Immonen S, Kalendar R (2000) Genome evolution of wild barley (Hordeum spontaneum) by BARE-1 retrotransposon dynamics in response to sharp microclimatic divergence. Proc Natl Acad Sci U S A. **97**:6603-7.

Shepherd NS, Schwarz-Sommer Z, Blumberg vel Spalve J, Gupta M, Wienand U, Saedler H (1984) Similarity of the Cin1 repetitive family of Zea mays to eukaryotic transposable elements. Nature. **307**:185-7.

Singer TC, Yordan C AND Martienssen R (2001) Robertson's *Mutator* transposons in *Arabidopsis thaliana* are regulated by the chromatin-remodeling gene, decrease in DNA methylation. Gene Dev. **15**:591-602.

Sinkovics JG (2001) The place of viruses in the "tree of life". Acta Microbiol Immunol Hung. **48**:115-27.

Staginnus C, Winter P, Desel C, Schmidt T, Kahl G (1999) Molecular structure and chromosomal localization of major repetitive DNA families in the chickpea (Cicer arietinum L.) genome. Plant Mol Biol. **39**:1037-50.

Steimer A, Amedeo P, Afsar K, Fransz P, Scheid OM, Paszkowski J (2000) Endogenous targets of transcriptional gene silencing in Arabidopsis. Plant Cell. **12**:1165-78.

Suoniemi A, Narvanto A, Schulman AH (1996) The BARE-1 retrotransposon is transcribed in barley from an LTR promoter active in transient assays. Plant Mol Biol. **31**:295-306.

Takeda S, Sugimoto K, Kakutani T, Hirochika H (2001) Linear DNA intermediates of the *Tto*1 retrotransposon in Gag particles accumulated in stressed tobacco and Arabidopsis thaliana. Plant J. **28**:307-17.

Takeda S, Sugimoto K, Otsuki H and Hirochika H (1999) A 13-bp cis-regulatory element in the LTR promoter of the tobacco retrotransposon *Tto1* is involved in responsiveness to tissue culture, wounding, methyl jasmonate and fungal elicitors. Plant J. **18:** 383–393.

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

Tatout C, Lavie L, Deragon JM (1998) Similar target site selection occurs in integration of plant and mammalian retroposons. J Mol Evol. **47**:463-70.

Temin HM (1980) Origin of retroviruses from cellular moveable genetic elements. Cell. **21**:599-600. Review.

Temin HM, Mizutani S (1970) RNA-dependent DNA polymerase in virions of Rous sarcoma virus. Nature. **226**:1211-3.

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.

Vaucheret H, Beclin C, Fagard M (2001) Post-transcriptional gene silencing in plants.J Cell Sci. **114**(Pt 17):3083-91. Review.

Vaucheret H, Beclin C, Elmayan T, Fuerbacj C, odon J.-B, Mourrain P, Plauqui J.-C, Vernhettes S (1998) Transgene-induced gene silencing in plants. Plant J. **16:** 651-659.

Vernhettes S, Grandbastien MA, Casacuberta JM (1998) The evolutionary analysis of the Tnt1 retrotransposon in Nicotiana species reveals the high variability of its regulatory sequences. Mol Biol Evol. **15**:827-36.

Vershinin AV, Druka A, Alkhimova AG, Kleinhofs A, Heslop-Harrison JS (2002) LINEs and gypsy-like retrotransposons in Hordeum species. Plant Mol Biol. **49**:1-14.

Vicient CM, Jaaskelainen MJ, Kalendar R, Schulman AH (2001) Active retrotransposons are a common feature of grass genomes. Plant Physiol. **125**:1283-92.

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.

Vielle-Calzada J.-P, Baskar A and Grossniklaus U (2000) Delayed activation of the paternal genome during seed development. Nature.**4**: 91–94.

Voytas DF, Cummings MP, Konieczny A, Ausubel FM, Rodermel SR (1992) Copia-like retrotransposons are ubiquitous among plants. PNAS USA **89:**7124-7128.

Voytas DF and Ausubel FM (1988) A copia-like transposable element family in Arabidopsis thaliana. Nature. **366:** 242-244.

Walbot V (2000) Saturation mutagenesis using maize transposons. Curr. Opin. Plant Biol. **3:** 103-107.

Walbot V and Cullis CA (1985) Rapid genomic change in higher plants. Annu Rev Plant Physol. **36**: 367-396.

Wang SP, Liu N, Peng KM, Zhang QF (1999) The distribution and copy number of *copia* in rice (*Oryza sativa*) and their implications in the organization and evolution of the rice genome. Proc. Natl. Acad. Sci. USA. **91**:11792-11796.

White SE, Bureau TE, Wessler SE (1994) Transduction of a cellular gene by a plant retroelement.Cell. 77(4):479-80.

Wessler SR, Feschotte C (2001) Treasures in the attic: rolling circle transposons discovered in eukaryotic genomes.Proc Natl Acad Sci US A. **98**:8923-4.

Wessler SR (1998) Phenotypic diversity mediated by the maize transposable elements *Ac* and *Spm*. Science **242:** 399–405.

Wessler SR (1996) Turned on by stress. Plant retrotransposons.Curr Biol. **6**:959-61. Review.

Wessler SR, Bureau TE, White SE (1995) LTR-retrotransposons and MITEs: important players in the evolution of plant genomes.Curr Opin Genet Dev. **5**:814-21.

Wessler SR, White SE, Habera LF (1994) Retrotransposons in the flanking regions of normal plant genes a-role for copia-like elements in the evolution of gene structure and expression. PNAS, USA. **91:** 11792-11796.

Wessler SR, Bureau TE (1992) Tourist: a large family of small inverted repeat elements frequently associated with maize genes.Plant Cell. **4**:1283-94.

Wicker T, Stein N, Albar L, Feuillet C, Schlagenhauf E, Keller B (2001) Analysis of a contiguous 211 kb sequence in diploid wheat (Triticum monococcum L.) reveals multiple mechanisms of genome evolution.Plant J. **26**:307-16.

Witte CP, Le QH, Bureau T, Kumar A (2001)Terminal-repeat retrotransposons in miniature (TRIM) are involved in restructuring plant genomes. Proc Natl Acad Sci U S A. **98**:13778-83.

Wright DA, Voytas DF (2002) Athila4 of Arabidopsis and Calypso of soybean define a lineage of endogenous plant retroviruses. Genome Res. **12**:122-31.

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

Yoshioka Y, Matsumoto S, Kojima S, Ohshima K, Okada N, Machida Y (1993) Molecular characterization of a short interspersed repetitive element from tobacco that exhibits sequence homology to specific tRNAs. Proc Natl Acad Sci U S A. **90**:6562-6.

Yu GX, Wise RP (2000) An anchored AFLP- and retrotransposon-based map of diploid Avena. Genome. **43**:736-49.

List of Accession Numbers

CART 155	AJ534846	CART 140	AJ535880
CART 5	AJ535748	CART 141	81
CART 6	AJ535749	CART 145	82
CART 9	AJ535750	CART 157	83
13	AJ535751	CART 186	84
24	AJ535855	CART 204	85
33	AJ535856	CART 210	86
37	57	CART 241	87
40	58	CART 273	88
50	59	CART 311	89
52	60	CART 324	90
54	61	CART 329	91
55	62	CART 345	92
56	63	CART 364	93
61	64	CART 366	94
67	65		
73	66	CARE 2	AJ544276
77	67	CARE 3	77
81	68	CARE 4	78
97	69	CARE 5	79
98	70	CARE 7	80
103	71		
105	72		
114	73		
120	74		
123	AJ535877		
126	78		
135	79		