Mol Gen Genomics (2006) 276:254–263
DOI 10.1007/s00438-006-0140-x

ORIGINAL PAPER

Retand: a novel family of gypsy-like retrotransposons harboring an amplified tandem repeat

Eduard Kejnovsky · Zdenek Kubat · Jiri Macas · Roman Hobza · Jaroslav Mracek · Boris Vyskot

Received: 20 December 2005 / Accepted: 2 June 2006 / Published online: 7 July 2006

Abstract In this paper we describe a pair of novel Ty3/gypsy retrotransposons isolated from the dioecious plant *Silene latifolia*, consisting of a non-autonomous element *Retand-1* (3.7 kb) and its autonomous partner *Retand-2* (11.1 kb). These two elements have highly similar long terminal repeat (LTR) sequences but differ in the presence of the typical retroelement coding regions (*gag-pol* genes), most of which are missing in *Retand-1*. Moreover, *Retand-2* contains two additional open reading frames in antisense orientation localized between the *pol* gene and right LTR. *Retand* transcripts were detected in all organs tested (leaves, flower buds and roots) which, together with the high sequence similarity of LTRs in individual elements, indicates their recent transpositional activity. The autonomous elements are similarly abundant (2,700 copies) as non-autonomous ones (2,100 copies) in *S. latifolia* genome. *Retand* elements are also present in other *Silene* species, mostly in subtelomeric heterochromatin regions of all chromosomes. The only exception is the subtelomere of the short arm of the Y chromosome in *S. latifolia* which is known to lack the terminal heterochromatin. An interesting feature of the *Retand* elements is the presence of a tandem repeat sequence, which is more amplified in the non-autonomous *Retand-1*.

Keywords Pair of autonomous and non-autonomous retrotransposons · Tandem repeat · Transcriptional activity · Sex chromosomes · *Silene latifolia*

Communicated by M.-A. Grandbastien

Nucleotide sequence data reported are available in the DDBJ/EMBL/GenBank databases under the accession numbers DQ023669 (*Retand-2*) and DQ023670 (*Retand-1*).

E. Kejnovsky (✉) · Z. Kubat · R. Hobza · J. Mracek · B. Vyskot
Laboratory of Plant Developmental Genetics,
Institute of Biophysics,
Academy of Sciences of the Czech Republic,
Kralovopolska 135, 612 65 Brno, Czech Republic
e-mail: kejnovsk@ibp.cz

J. Macas
Laboratory of Molecular Cytogenetics,
Institute of Plant Molecular Biology,
Academy of Sciences of the Czech Republic,
37005 Ceske Budejovice, Czech Republic

Introduction

Retrotransposons represent a class of mobile elements with duplicative mode of transposition via RNA intermediate. They are widespread among eukaryotes and often compose a significant fraction of the nuclear genome (Bennetzen 1996; SanMiguel et al. 1996). In plants, retrotransposons are found in all genomes that have been examined, but they seem to be highly abundant only in species with large genomes, in some plants comprising more than 50% of nuclear DNA content (Pearce et al.1996a; SanMiguel et al. 1996; Suoniemi et al. 1998). The replication strategy of retrotransposons offers the potential for explosive increases in copy numbers and thereby these elements increase the size of the host genome (Kumar and Bennetzen 1999; SanMiguel and Bennetzen 1998). Differential amplification of retrotransposons is largely responsible for significant differences in genome size even between closely related plant species (SanMiguel 1998; Vicient et al. 1999; Kalendar et al. 2000; SanMiguel and Bennetzen 1998).

It has been found that plant genomes host diverse populations of retroelements which are classified into several basic classes (Hull 2001), each of them is represented by various families including elements similar in their nucleotide sequences and thus supposed to be of the same origin. However, even the retroelement sequences belonging to the same family may differ considerably in their length and structure, as they occur in the genome as deletion derivatives of the full-length retroelements (Havecker et al. 2004). A typical example is solo-LTRs derived from LTR-retrotransposons by recombination-based excision and elimination of internal element sequences between the two terminal LTRs. In addition, non-autonomous elements lacking a part or whole of the regions coding for retroelement proteins have been reported from several species. As these elements are not capable of encoding functional proteins required for their replication and/or transposition, they can be propagated in the genome only at the presence of a functional intact element. However, the molecular mechanisms producing these non-autonomous elements, as well as the population dynamics of autonomous and non-autonomous elements in plant genomes are only poorly understood. Moreover, except for a few extensively sequenced model species like rice (Jiang et al. 2002a, b), only a little is known about the occurrence of these elements in other plants.

Here we describe *Retand*, a novel family of Ty3/gypsy retrotransposons including the full-length and the corresponding non-autonomous elements occurring in the genome of *Silene latifolia*. These elements are localized at subtelomeric heterochromatin regions of a vast majority of chromosomes and are transcriptionally active. An interesting feature of these retrotransposons is the presence of an array of tandem repeats within their sequences which is more amplified in the non-autonomous element, as well as the presence of two extra ORFs in antisense orientation located upstream of right LTR of the autonomous element.

Materials and methods

Cloning and hybridization

The construction of a sample BAC library of the male *S. latifolia* and the preparation of a membrane with 960 BAC clones were described in Lengerova et al. (2004). Screening was performed by hybridization with AlkPhos Direct hybridization kit (Amersham Pharmacia Biotech) according to the manufacturer's protocol. Probes were prepared by PCR amplification of LTR region, regions of *gag* and *pol* genes, ORF3 region and STR1 tandem repeat region with following primers, respectively: LTR-F1 (5′-TTCCGGGTGTAATTCC AGAG-3′), LTR-R1 (5′-CATATTCTGCACCCGCT GAC-3′), GAG-F1 primer (5′-CCGACTCTCCAAA GTCGAAC-3′) and GAG-R1 (5′-TTCTGCCCTTC ATTCTTGCT-3′), POL-F1 (5′-ACGGTCAGTGCC GTAATCATC-3′), POL-R1 (5′-GTCGTCTCGGCT TCTATGTC-3′), STR1-F1 (5′-GAGGTGTCCAAG ACCGTTGT-3′), STR1-R1 (5′-GCCGTCGACTTCT TTCTTGT-3′), ORF4-F1 (5′-TAACGTAACCCGGC TCTGTC-3′) and ORF4-R1 (5′-TTTGGAGGATTTC GAGGATG-3′) with BAC4J4 clone containing an autonomous *Retand* element as a template DNA. After hybridization, detection was performed using CDP-Star detection reagent. BAC DNA was isolated using the Qiagen Plasmid Midi kit. BAC DNA was digested with restriction enzymes and desired fragments containing parts of *Retand* elements were subcloned into linearized and dephosphorylated plasmid pBluescript IISK$^+$ and transformed into XL1-Blue cells. Subcloned fragments were sequenced using GeneJumper Primer Insertion Kit for Sequencing (Invitrogen) based on transposon-mediated insertions of primer binding sites. All parts of the *Retand* elements were read 2–4 times (by VBC-Genomics, Vienna) and the clone sequences were assembled with Staden Package software (Staden et al. 1996).

Fluorescence in situ hybridization

In order to synchronize the germinating seeds of *S. latifolia* and *S. viscosa* (seeds come from the seed collection of the Institute of Biophysics, Brno, Czech Republic), the DNA polymerase inhibitor aphidicolin was added for 12 h, and mitoses were then accumulated with oryzalin as described in Siroky et al. (2001). Slides were treated as described in Lengerova et al. (2004) with slight modifications. Slide denaturation was performed in 7:3 (v/v) formamide: 2× SSC for 2 min at 72°C. Slides were immediately dehydrated through 50, 70, and 100% ethanol (−20°C), and air dried. The probe was denatured at 70°C for 10 min, and 100 ng of the denatured probe was added at room temperature and hybridized for 18 h at 37°C. Slides were analyzed using Olympus Provis microscope, and image analysis was performed using ISIS software (Metasystems). DNA was labeled with Fluorolink Cy3-dUTP (Amersham Pharmacia Biotech) (red labeling) in combination with the nick translation mix (Roche) or with SpectrumGreen direct-labeled dUTP (green labeling) and the Nick Translation kit (both Vysis).

RNA isolation and RT-PCR

Total RNA was isolated from *S. latifolia* male plant tissues (leaves, flower buds or roots) by extraction with RNA Blue (TopBio) and chloroform. All RNA samples were treated with RNase free DNase (Gibco-BRL) to remove any contaminant DNA. Reverse transcription was carried out with the SuperScriptIII First-Strand Synthesis System for RT-PCR (Gibco-BRL) by random priming method, with 1 μg of the template RNA. RNase H was added to remove the RNA template from the cDNA:RNA hybrid molecule after first-strand synthesis. The RT-PCR reaction mix (50 μl) consisted of 1× PCR buffer, 0.2 mM dNTP, 0.2 μM primers, 1.5 mM MgCl2, 1 U of Taq polymerase (TopBio) and 10–100 ng of reverse-transcribed RNA or equal amount of reverse transcriptase-untreated RNA as a negative control. The following primers were used in the RT-PCR experiments (sequences of primers see above): LTR-F1, LTR-R1, POL-F1, POL-R1, STR1-F1, STR1-R1, ORF3-F1 (5′-CCATTAGTGGGCGACCTAGA-3′), ORF3-R1 (5′-TTCAGGCTGCCTACGACTTT-3′), ORF4-F1, ORF4-R1, Actin-F1 (5′-CCTTGTCTGTGACAATGG AAC-3′) and Actin-R1 (5′- GCTCACAATACCGTG CTCAA-3′). The reaction profile included 35 cycles of 40 s at 94°C, 1 min at 55°C and 1 min at 72°C preceded by initial denaturation (2 min at 94°C) and followed by final extension step (7 min at 72°C).

DNA sequence analysis

Basic sequence analysis was done with Staden Package software (Staden et al. 1996). Multiple sequence comparisons were performed with ClustalW (Thompson et al. 1994) and plots of sequences were done with JDotter (Brodie et al. 2004; http://www.athena.bioc.uvic.ca/pbr/jdotter/). Homology search was performed with BLAST and FASTA (Altshul et al. 1997; and Pearson and Lipman 1988, respectively). The search for conserved protein domains was done with RPS-Blast (Marchler-Bauer et al. 2003). Sequences of tRNA used for identification of PBS were obtained from the *Arabidopsis thaliana* tRNA database (Lowe and Eddy 1997; http://www.lowelab.ucsc.edu/GtRNAdb/).

Results

Structure of non-autonomous *Retand-1* and its autonomous partner *Retand-2* in *S. latifolia*

In the course of characterization of the BAC4G8 clone giving prominent subtelomeric fluorescence in situ hybridization (FISH) signals, we found a novel non-autonomous retroelement that we named *Retand-1* because it contained a tandem repeat array instead of genes (retroelement with tandem repeat). This element was 3,679 bp long and contained almost identical long terminal repeats (LTRs) differing in only one base at 3′end (99.8% identity) indicating a recent insertion of the element. The LTRs began with TG and ended with CA, similarly as the LTRs of most plant retrotransposons. We propose that the sequence TATA-AAT located in the center of left LTR could serve as putative transcription initiation signal for *Retand-1*. At both ends of the element, we found 5 bp direct repeats CTTTG, which probably represent the target site duplications (TSD) generated during the element integration. Several features necessary for retroelement replication were also identified, including primer binding site (PBS) located downstream of the left LTR and polypurine tract (PPT) located upstream of the right LTR (Fig. 1). PBS was separated from the left LTR by 5 bp and showed complete identity to the 15 bp at the 3′ end of Arabidopsis $tRNA^{Lysine}$. The interesting feature of *Retand-1* was an array of tandem repeats, named STR1 (*Silene* tandem repeat 1), composed of 12 monomers located in the central part of element. Total length of tandem repeat array was 799 bp. Most monomers of STR1 represented perfect repeats 67 bp long, some contained one or few base substitutions, one monomer had 9 bp deletion and other monomer had 4 bp insertion (Fig. 1b). No significant homology has been found between STR1 and any other tandem repeat described to date.

In order to identify a full-length (master) *Retand* element, we screened the sample BAC library of *S. latifolia* with LTR of the *Retand-1* as the probe. Out of 960, 157 BAC clones hybridized with this probe. BAC4J4 clone containing a large restriction fragment hybridizing with LTR was selected, subcloned into a plasmid vector, and sequenced. The sequence analysis revealed the full-length element with a total length of 11,073 bp. This autonomous element, named *Retand-2*, in contrast to the smaller non-autonomous element *Retand-1*, contained all genes typical for retroelements. On the other hand, the array of tandem repeat STR1 was less amplified in *Retand-2* compared to *Retand-1*. *Retand-2* has two LTRs of 585 and 584 bp on left and right ends, respectively, with 98.5% identity. The sequence analysis of *Retand-2* revealed the presence of four ORFs (Fig. 2). Searching the database of conserved domains (RPS Blast) with the *Retand-2* sequence as a query allowed the identification of the region potentially coding for gag protein (the first ORF), prote-

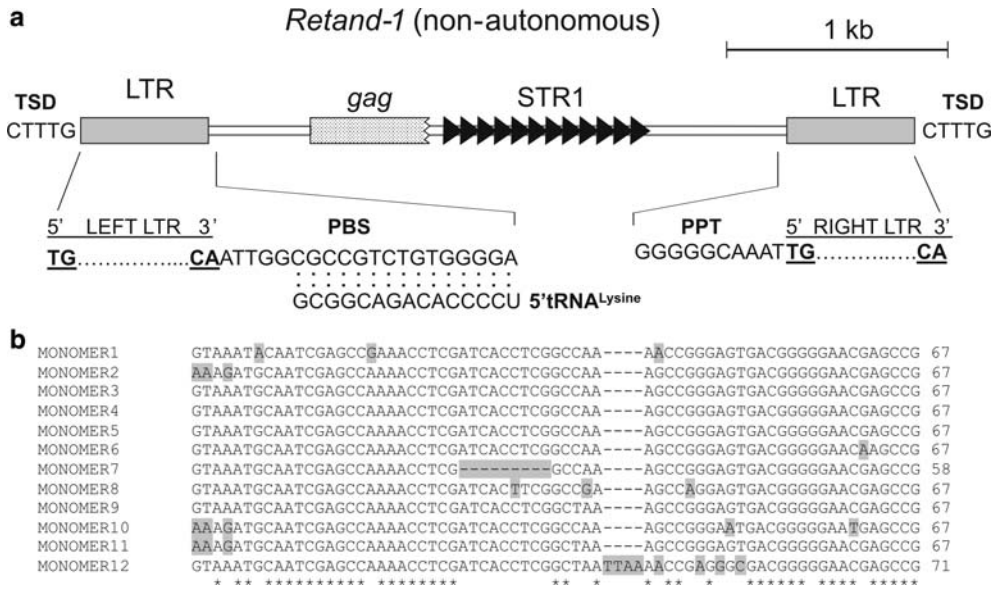

Fig. 1 **a** Scheme of the non-autonomous *Retand-1* element with structural and functional features marked as follows: *LTR* long terminal repeat (starting with TG and ending with CA); *PBS* primer binding site (its complementarity to the 3′ end of $tRNA^{Lysine}$ is indicated); *PPT* polypurine tract; *TSD* target site duplication. *Gag* gene is 3′end truncated. Tandem repeats STR1 are marked by *arrows* each corresponding to one monomer. **b** Multiple alignments (by ClustalW) of 12 monomers of tandem repeat STR1. Missing nucleotides are represented by *dash*; nucleotides that differ from prevalent bases are *shaded*; *asterisks* indicate positions where nucleotides are identical in all monomers. The *bar* equal to the length of 1 kb is indicated

ase, reverse transcriptase/RNase H and integrase (second ORF corresponding to *pol* gene). The order of the coding domains indicated that *Retand-2* could be classified as LTR retrotransposon of the Ty3/gypsy type (the family of Metaviridae, Hull 2001). Downstream of the *pol* gene there is an array of tandem repeats with seven monomers homologous to the STR1 tandem repeat of *Retand-1* element. The remaining two ORFs, named ORF3 and ORF4, are in antisense orientation. They are located between the STR1 tandem repeat and the right LTR; these ORFs are 896 and 1,676 bp long, and overlap by 28 bp. The search for conserved protein domains done with RPS-Blast revealed the homology of ORF4 to the transposase 28 domain of gypsy-type retrotransposon (pfam04195, E-value: 3e−05), thus probably representing the remnant of other mobile element inserted into *Retand-2*. Comparison of the autonomous *Retand-2* and non-autonomous *Retand-1* elements showed their homology in LTRs (left LTRs showed 85% identity, right LTRs showed 86% identity) as well as in the region downstream of the left LTR where *gag* gene is located. Similarly, there is a homology between both *Retand* elements in the region upstream of the right LTR. The *gag* gene is 3′ end truncated in *Retand-1*.

In order to determine the lengths of STR1 tandem repeat arrays in the genome of *S. latifolia*, we partially digested genomic DNA with restrictases *Taq*I or *Mbo*I, cleaving within the STR1 monomer, and hybridized Southern blot with STR1 as a probe (Fig. 3). We observed a ladder-like pattern in which bands corresponded to monomer, dimer, trimer, etc., reaching the size of the arrays up to several kilobases. The variability in the length of STR1 arrays inside various autonomous *Retand* elements were determined by PCR in ten selected BAC clones containing *Retand*. PCR with primers complementary to the sequence flanking the STR1 array in *Retand-2* resulted in products in nine of ten BAC clones; their length was either the same as in the *Retand-2* or differed at maximum 200 bp which corresponded to maximum three monomers.

Retand is ubiquitously transcribed

The transcriptional activity of the *Retand* elements was studied by RT-PCR with total RNA extracted from leaves, flower buds, or roots as a template. The primers were designed to amplify the regions of LTR, *pol* gene, STR1 tandem repeat, ORF3 or ORF4. Especially primers derived from the LTR region, which is the most variable among retrotransposons, ensured that expression of *Retand* sequences will be specifically followed. Two sets of otherwise identical reactions differing only in including or omitting the

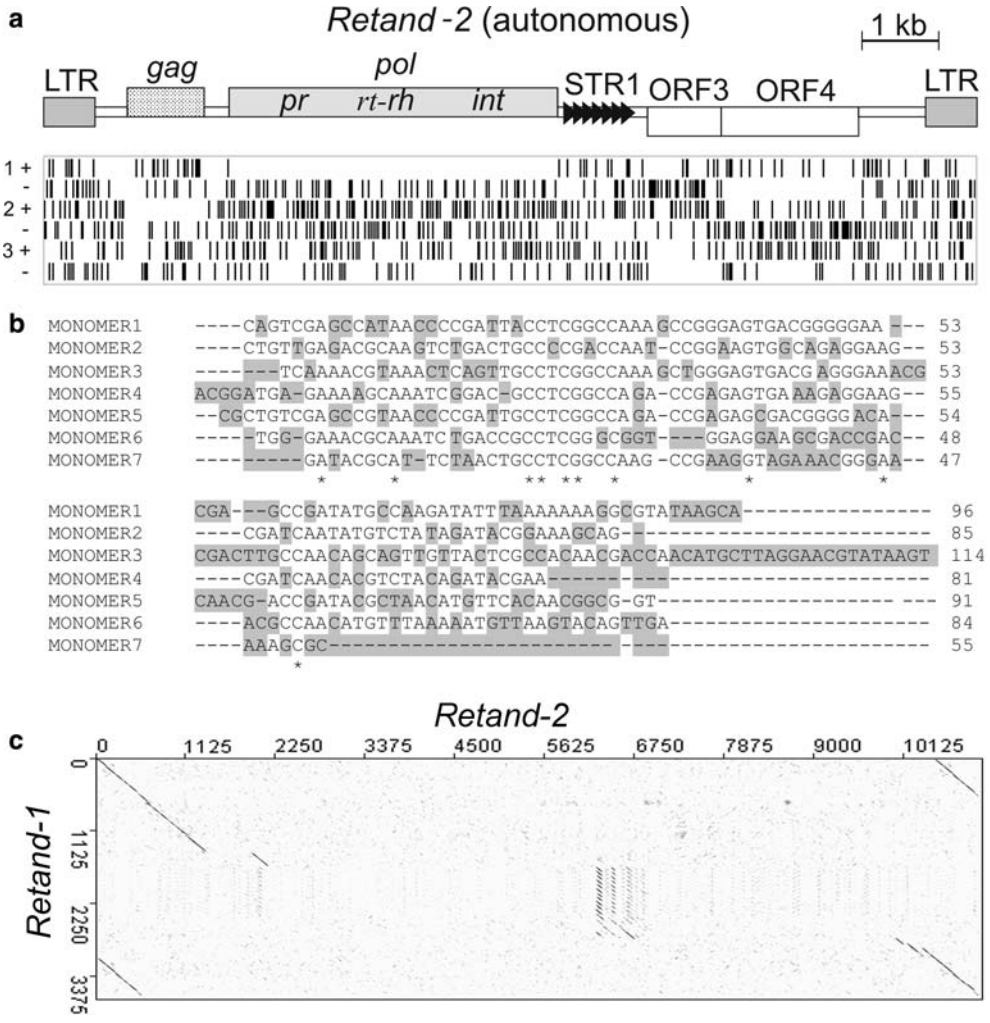

Fig. 2 **a** Scheme of the autonomous *Retand-2* element with indicated LTRs (*dark-shaded boxes*), *gag* gene (*spotted box*), and *pol* gene (*light-shaded box*) with positions of individual coding domains (pr, rt-rh, int). Two ORFs downstream of tandem repeats STR1 (*arrows*), ORF3 and ORF4, are marked by *open boxes*. Positions of stop codons are plotted in three sense and three antisense reading frames. The *bar* equal to the length of 1 kb is indicated. **b** Multiple alignments (by ClustalW) of seven monomers of tandem repeat STR1. Missing nucleotides are represented by *dash*; nucleotides that differ from prevalent bases are *shaded*; *asterisks* indicate positions where nucleotides are identical in all monomers. **c** A dot matrix comparison (by JDotter) of autonomous *Retand-2* (horizontally) versus non-autonomous *Retand-1* (vertically) elements shows the homology in LTRs, a part of *gag* gene, and the tandem repeats STR1. Homologous sequence regions are represented by *dots* or *lines*

reverse transcription step were always run in order to check eventual false-positive results caused by contamination with genomic DNA (Fig. 4). In RT-PCR experiments with LTR, *pol* gene and STR1 primers fragments of expected size were amplified. The LTR and *pol* gene were equally amplified from leaves, flower buds and roots; however, the expression of STR1 was reduced in flower buds. In ORF3 and ORF4, located in antisense orientation, only weak expression was detected in all organs tested. Primers amplifying a part of actin gene, which is always spliced after transcription, were used as a control and resulted in longer band in genomic DNA template (non-spliced copy) and shorter band (spliced) when RNA was used as a template. These data demonstrate that *Retand* is expressed in all organs tested—leaves, flower buds and roots. We have cloned and sequenced the part of *pol* gene amplified by RT-PCR in different organs (six to nine clones from each organ—leaf, flower bud or root) to answer the question whether only one or more *Retand* elements are active. We found high polymorphism in *pol* transcripts indicating that many different *Retand* elements are transcribed. Most of the *pol* gene transcripts (83%) did not contain any stop codons and thus could be translated. Comparison of sequences from different organs showed that there were no organ-specific fragments.

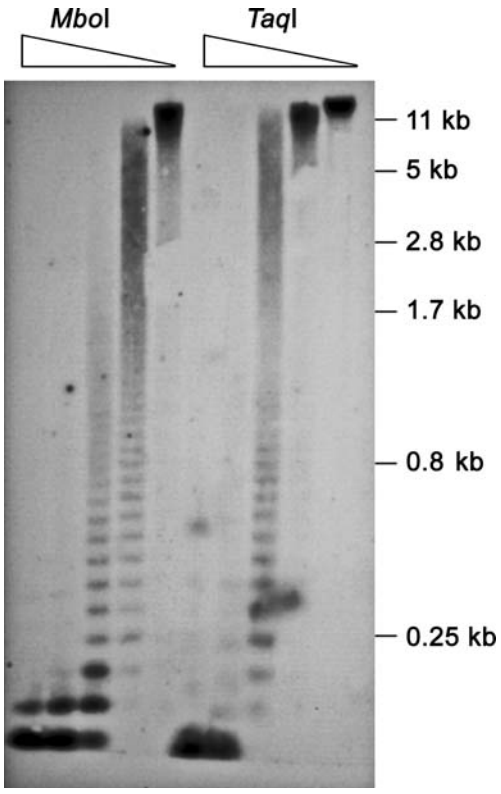

Fig. 3 Southern analysis of *S. latifolia* male genomic DNA partially digested with decreasing amounts of restrictases *Mbo*I or *Taq*I (having restriction sites within the STR1 tandem repeat monomer) and hybridized with the STR1 tandem repeat as a probe

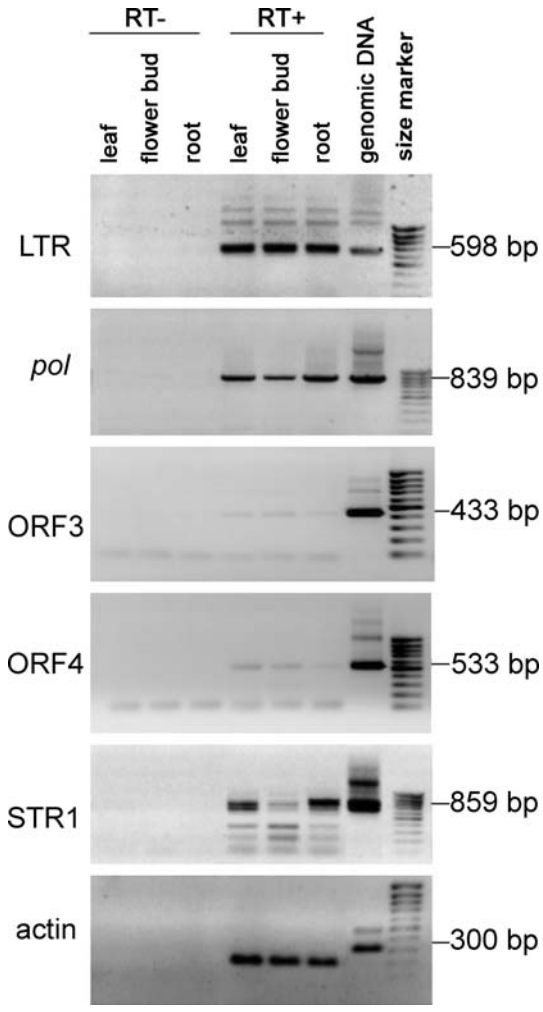

Fig. 4 Transcriptional activity of *Retand* analyzed by RT-PCR. The primers were directed to amplify the following portions of *Retand* transcripts: long terminal repeat (LTR), *pol* gene (*pol*), ORF3, ORF4 and tandem repeat STR1 (STR1). The reactions were performed parallel with total RNA isolated from leaves, flower buds or roots (RT-) and with the adequate amounts of reversely transcribed RNA (RT+) into single strand cDNA. Primers amplifying housekeeping actin gene, which is spliced, were used as a control. Actin primers amplify non-spliced copy on genomic DNA template

Retand is abundant in *S. latifolia* genome and is also present in other *Silene* species

The abundance of the *Retand* elements was determined by the screening of a sample BAC library with probes derived from defined regions of the element. Four different probes, representing LTR, *gag* gene, *pol* gene, or STR1 tandem repeat regions, were separately used for hybridization. Screening of the sample BAC library resulted in 16% of positive clones for LTR (157 of 960), 13% of positive clones for *gag* (129 of 960), 9.4% positive clones for *pol* gene (90 of 960), 7.6% positive clones for STR1 tandem repeat region (73 of 960), and 8.2% of ORF4 (79 of 960). Taking into account the *S. latifolia* male genome size (1C = 2,925 Mb, Siroky et al. 2001) and the average insert size in our BAC library (100 kb), the estimated numbers of copies per haploid genome were ca. 4,800 determined by hybridization with LTR, 3,900 for *gag*, 2,700 for *pol*, 2,200 for tandem repeat region, and 2,400 for ORF4. This calculation is based on presumption that each positive clone contained only a single copy of the element. To test this, ten selected BAC clones digested separately with four restrictases were hybridized with LTR probe to find that in most clones (nine of ten) maximum three bands contained parts of LTR as a result of restrictase cleavage inside LTR and/or between two LTRs of single element (data not shown). Ninety BAC clones contained *pol* gene and 79 of them contained also ORF4 indicating that most (88%) autonomous *Retand* elements contained ORF4. It suggested that *Retand-2* characterized here is a typical autonomous *Retand* element. These data indicate that (1) there is a similar number of non-autonomous and autonomous *Retand* elements (2,100 and 2,700 copies, respectively), and (2) tandem repeat STR1 is present in about a half of *Retand* elements (2,200 of 4,800 elements).

The presence of *Retand* in other related *Silene* species, *S. dioica*, *S. diclinis*, *S. vulgaris* and *S. viscosa* was studied by genomic Southern blotting. Genomic DNA was digested with different restrictases and hybridized with LTR as a probe. Strong signals in all tested genomic DNAs demonstrated that *Retand* was present in high copy numbers in all the tested *Silene* species (data not shown).

Retand is localized at subtelomeres in *Silene* species

In order to study chromosomal localization of the *Retand* elements, we used LTR as a probe for FISH. In *S. latifolia*, strong subtelomeric signals of similar intensities in a vast majority of chromosomes were observed. The only one exception was the Y chromosome having signal only on the q arm (Fig. 5a) containing the pseudoautosomal region (Lengerova et al. 2003) and hybridizing with a marker subtelomeric repetitive sequence, X43.1 (Fig. 5b) described by Buzek et al. (1997). FISH with probe derived from the *pol* gene reflecting the localization of full-length elements resulted in a similar pattern compared to LTR probe. The intensity of signal localized on the Y chromosome was significantly lower compared to the X chromosome and autosomes (Fig. 5c).

Fluorescence in situ hybridization performed with the STR1 tandem repeat as a probe gave much weaker signals of variable intensities located at subtelomeres of most chromosomes (Fig. 5d). This result is consistent with the presence of tandem arrays in about half of the *Retand* elements. FISH revealed similar subtelomeric localization both in closely related species, *S. diclinis* and *S. dioica* (data not shown), and in a more distantly related *S. viscosa* (Fig. 5e).

Discussion

On the basis of the order of coding domains, *Retand* can be classified as a Ty3/gypsy-type retrotransposon (Hull 2001). However, in the autonomous element of *Retand-2*, there are two additional ORFs in antisense orientation at the 3′ end which are present in most *Retand* elements. One of them, ORF4, has a homology to transposase 28 domain and is probably a remnant of a mobile element inserted into *Retand* in the past which lost its other parts. Recently, a number of retrotransposon families have been identified, some of them have also additional coding information. One group represents retrotransposons with "env-like" genes (Peterson-Burch et al. 2000; Zaki 2003; Neumann et al. 2005, and references therein). Other novel coding regions have been identified within various retrotransposons, e.g. *Cyclops* (Chavanne et al. 1998) and *Ogre* (Neumann et al. 2003). Similarly to *Retand-2*, the additional ORFs present in *RIRE2* of rice (Ohtsubo et al. 1999) and *Grande1* of maize (Martinez-Izquierdo et al. 1997)

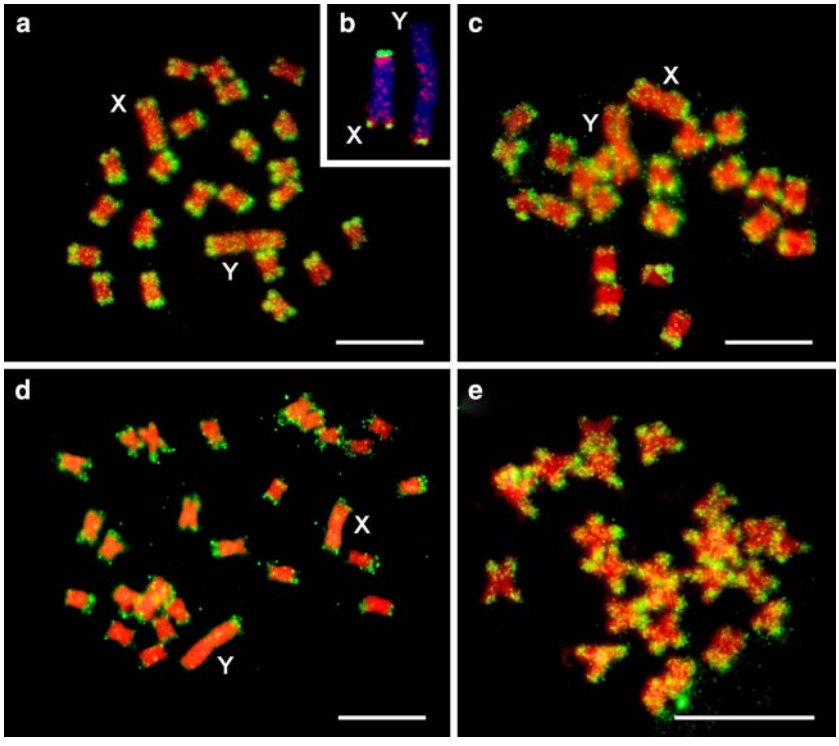

Fig. 5 Chromosomal distribution of the *Retand* sequences analyzed by FISH. Metaphase chromosomes of *S. latifolia* (**a**–**d**), and *S. viscosa* (**e**) were hybridized with LTR (**a**, **b**, **e**), *pol* sequence (**c**) or the tandem repeat STR1 (**d**). Chromosomes were counterstained with DAPI (*red*); probes were labeled with Cy3- or SpectrumGreen-conjugated nucleotides (here all signals are in *green* in **a**, **c**–**e**). **b** Bicolor FISH of sex chromosomes of *S. latifolia* hybridized with LTR (*red*) and X43.1 repeat (*green*); DAPI is in *blue* for contrast. The X and Y chromosomes in *S. latifolia* are marked, *bars* = 10 μm

are also in antisense orientation. These ORFs have also similar position upstream of the right LTR as was found in *Retand-2*. The function, if any, of antisense ORFs is unknown.

The presence of the non-autonomous element *Retand-1* is another remarkable feature of the *Retand* family. This phenomenon is well documented in DNA transposons (Feschotte et al. 2002) but recent genome-mining studies revealed also non-autonomous retrotransposons demonstrating their high diversity (Havecker et al. 2004). *Retand-1*, similarly as other non-autonomous retroelements, contains only minimal sequences required for replication including LTRs, a primer binding site and PPT. The best known examples of a pair of autonomous and non-autonomous elements are *RIRE2* and *Dasheng* from rice (Jiang et al. 2002b). Large non-autonomous elements like *Dasheng*, ranging in size from 5.5 to 8.6 kb, have now been named large retrotransposon derivates (LARDs) (Kalendar et al. 2004). A non-autonomous *Retand-1* element with size of 3.7 kb could be also assigned to this group. *Retand* family with about 5,000 copies, about half of which are non-autonomous elements, is another example demonstrating staggering success of non-autonomous elements. The high identity of LTRs of *Retand-1* (99.8% identity, differing in 1 bp over 608 bp) as well as of *Retand-2* (98% identity of LTRs) demonstrates their recent insertion and suggests that *Retand* is an active element. It is also supported by its observed ubiquitous transcriptional activity.

Most of the *Retand* elements are localized in subtelomeres of most chromosomes. The absence of subtelomeric heterochromatin at the p arm of the Y chromosome revealed by DAPI banding (Buzek et al. 1997), where *Retand* is missing (Fig. 5a, b), indicates that *Retand* is concentrated at heterochromatic regions. Thus, it represents another example demonstrating that subtelomeric or telomeric regions appear to be particularly favorable refuges for retroelements (Pearce et al. 1996b). In some cases retroelement adaptation to these niches may have been so efficient that it resulted in the replacement of conventional telomerase activity (Pardue et al. 1996). In cereals, Ty3/gypsy retrotransposons have been accumulated in the centromeric regions as was shown in sorghum (Miller et al. 1998), rice (Dong et al. 1998), barley (Presting et al. 1998), maize (Ananiev et al. 1998), and wheat (Fukui et al. 2001). In *Arabidopsis*, retroelements are enriched in the pericentromeric regions rather than that in centromeres (Fransz et al. 2000) while in banana *monkey* retroelements are concentrated in the nucleolar organizer region (Balint-Kurti et al. 2000). Subtelomeric localization of *Retand* in *S. latifolia* is specific for this element because most of retroelements are distributed ubiquitously along all chromosomes as was shown by Matsunaga et al. (2002) when reverse transcriptase domain was used as a probe for FISH.

An interesting feature of *Retand* is, similarly to *Dasheng*, its array of STR1 tandem repeat which is less amplified in autonomous element *Retand-2* than in the non-autonomous element *Retand-1*. So far, the arrays of tandem repeats were found in several mobile elements. They are located upstream of the right LTR in maize retrotransposons *Grande1* (Martinez-Izquierdo et al. 1997) and *Cinful* (Sanz-Alferez et al. 2003), and *Drosophila* retrotransposon *micropia* (Lankenau et al. 1994). A special group represents retrotransposons containing the arrays of tandem repeats but lacking coding capacity. In addition to *Dasheng* in rice (Jiang et al. 2002a), it includes, e.g., *ZLRS* element from maize (Monfort et al. 1995). So far, no exact function has been attributed to the tandem repeats present in retrotransposones, although their conserved location among some Ty3/gypsy elements indicates their potential structural or functional significance (Martinez-Izquierdo et al. 1997; Sanz-Alferez et al. 2003; Lankenau et al. 1994). In order to test their potential to form secondary structures, we performed an analysis using Mfold software (Zuker 2003) for both *Retand-1* and *Retand-2* tandem repeat arrays as well as for some other tandem repeats located inside mobile elements, including micropia and *Grande1*. However, in all tested sequences the potential of tandem repeats to form secondary structures was comparable to random sequences under various ionic strengths and temperatures. Tandem repeats could serve as a junction between an internal domain containing *gag-pol* and its 3′ end part containing ORFs with unknown function in the location of retroviral *env* gene. A similar organization was recently observed in *Cinful* retrotransposon family (Sanz-Alferez et al. 2003).

The presence of tandem repeats within retroelement sequences raises the question about their eventual role in the origin of satellite repeats. Tandem repeats located inside mobile elements represent a pool of numerous, but relatively short tandemly organized regions, some of which might become amplified to longer arrays subjected to homogenization and amplification mechanisms acting on satellite repeats. Similar tandemly repeated regions occurring within rDNA intergenic spacers have already been proposed as a source of satellite repeats in plants (Unfried et al. 1991; Macas et al. 2003). However, in spite of several reports of highly amplified satellite repeats derived from other parts of retroelement sequences (Pasero et al. 1993; Rossi et al. 1993; Langdon 2000; Cheng and Murata

2003; Tek et al. 2005), none was shown to originate from a tandem repeat similar to that occurring in the *Retand* elements. FISH experiments on *S. latifolia* chromosomes described here also showed that there is no prominent satellite derived from the *Retand* repeats apart from original retrotransposon located at subtelomeres.

Acknowledgments This work was supported by the Czech Science Foundation (grant nos. 204/05/2097, 204/05/H505 and 204/04/1207) and AVOZ50510513 from the Academy of Sciences of the Czech Republic.

References

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

Ananiev EV, Phillips RL, Rines HW (1998) Complex structure of knob DNA chromosome 9: retrotransposon invasion into heterochromatin. Genetics 149:2025–2037

Balint-Kurti PJ, Clendennen SK, Dolezelova M, Valarik M, Dolezel J, Beetham PR, May GD (2000) Identification and chromosomal localization of the monkey retrotransposon in *Musa* sp. Mol Gen Genet 263:908–915

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

Brodie R, Roper RL, Upton C (2004) JDotter: a Java interface to multiple doplots generated by dotter. Bioinformatics 20:279–281

Buzek J, Koutnikova H, Houben A, Riha K, Janousek B, Siroky J, Grant S, Vyskot B (1997) Isolation and characterization of X chromosome-derived DNA sequences from a dioecious plant *Melandrium album*. Chromosome Res 5:57–65

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

Cheng ZJ, Murata M (2003) A centromeric tandem repeat family originating from a part of Ty3/gypsy-retroelement in wheat and its relatives. Genetics 164:665–672

Dong F, Miller JT, Jackson S, Wang G-L, Ronald PC, Jiang J (1998) Rice (Oryza sativa) centromeric regions consist of complex DNA. Proc Natl Acad Sci USA 95:8135–8140

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

Fransz PF, Armstrong A, de Jong JH, Parnell LD, van Drunen G, Dean C, Zabel P, Bisseling T, Jones GH (2000) Integrated cytogenetic map of chromosome arm 4S of *A. thaliana*: structural organization of heterochromatic knob and centromere region. Cell 100:367–376

Fukui K-N, Suzuki G, Lagudah ES, Rahman S, Appels R, Yamamoto M, Mukai Y (2001) Physical arrangement of retrotransopson-related repeats in centromeric region in wheat. Plant Cell Physiol 42:189–196

Havecker ER, Gao X, Voytas DF (2004) The diversity of LTR retrotransposons. Genome Biol 5:225

Hull R (2001) Classifying reverse transcribing elements: a proposal and a challenge to the ICTV. Arch Virol 146:2255–2261

Jiang N, Bao Z, Temnykh S, Cheng Z, Jiang J, Wing RA, McCouch SR, Wesler SR (2002a) Dasheng: a recently amplified nonautonomous long terminal repeat element that is a major component of pericentromeric regions in rice. Genetics 161:1293–1305

Jiang N, Jordan IK, Wessler SR (2002b) Dasheng and RIRE2. A nonautonomous long terminal repeat element and its putative autonomous partner in the rice genome. Plant Physiol 130:1697–1705

Kalendar R, Tanskanen J, Immonen S, Nevo E, 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, Vicient CM, Peleg O, Anamthawat-Jonsson K, Bolshoy A, Schulman AH (2004) Large retrotransposon derivates: abundant, conserved, but nonautonomous retroelements of barley and related genomes. Genetics 166:1437–1450

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

Langdon T, Seago C, Jones RN, Ougham H, Thomas H, Forster JW, Jenkins G (2000) De novo evolution of satellite DNA on the rye B chromosome. Genetics 154:869–884

Lankenau S, Corces VG, Lankenau DH (1994) The *Drosophila* micropia retrotransposon encodes a testis specific antisense RNA complementary to reverse transcriptase. Mol Cell Biol 14:1764–1775

Lengerova M, Moore RC, Grant SR, Vyskot B (2003) The sex chromosomes of *Silene latifolia* revisited and revised. Genetics 165:935–938

Lengerova M, Kejnovsky E, Hobza R, Macas J, Grant SR, Vyskot B (2004) Multicolor FISH mapping of the dioecious model plant, *Silene latifolia*. Theor Appl Genet 108:1193–1199

Lowe TM, Eddy SR (1997) tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. Nucleic Acids Res 25:955–964

Macas J, Navratilova A, Meszaros T (2003) Sequence subfamilies of satellite repeats related to rDNA intergenic spacer are differentially amplified on *Vicia sativa* chromosomes. Chromosoma 112:152–158

Marchler-Bauer A, Anderson JB, DeWeese-Scott C, Fedorova ND, Geer LY, He S, Hurwitz DI et al (2003) CDD: a curated Entrez database of conserved domain alignments. Nucleic Acids Res 31:383–387

Martinez-Izquierdo JA, Garcia-Martinez J, Vicient CM (1997) What makes Grande1 retrotransposon different? Genetica 100:15–28

Matsunaga S, Yagisawa F, Yamamoto M, Uchida W, Nakao S, Kawano S (2002) LTR retrotransposons in the dioecious plant *Silene latifolia*. Genome 45:745–751

Miller JT, Jackson SA, Nasuda S, Gill BS, Wing RA, Jiang J (1998) Cloning and characterization of a centromere-specific repetitive DNA element from *Sorghum bicolor*. Theor Appl Genet 96:832–839

Monfort A, Vicient CM, Raz R, Puigdomenech P, Martinez-Izquierdo J (1995). Molecular analysis of a putative transposable retroelement from the *Zea* genus with internal clusters of tandem repeats. DNA Res 2:255–261

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

Neumann P, Pozarkova D, Koblizkova A, Macas J (2005) PIGY, a new plant envelope-class LTR retrotransposon. Mol Gen Genomics 273:43–53

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

Pardue ML, Danilevskaya ON, Lowenhaupt K, Slot F, Traverse KL (1996) *Drosophila* telomeres: new view on chromosome evolution. Trend Genet 12:48–52

Pasero P, Sjakste N, Blettry C, Got C, Marilley M (1993) Long-range organization and sequence-directed curvature of *Xenopus laevis* satellite 1 DNA. Nucleic Acids Res 21:4703–4710

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

Pearce SR, Pich U, Harrison G, Heslop-Harrison JS, Flavell AJ, Schubert I, Kumar A (1996b) The Ty1-copia group retrotransposons of *Allium cepa* are distributed throughout the chromosomes but are enriched in the terminal heterochromatin. Chromosome Res 4:365–371

Pearson WR, Lipman DJ (1988) Improved tools for biological sequence comparison. Proc Natl Acad Sci USA 85:2444–2448

Peterson-Burch BD, Wright DA, Laten HM, Voytas DF (2000) Retroviruses in plants? Trends Genet 16:151–152

Presting GG, Malysheva L, Fuchs J, Schubert I (1998) A Ty3/gypsy retrotransposon-like sequences localize to the centromeric regions of cereal chromosomes. Plant J 16:721–728

Rossi MS, Pesce CG, Reig OA, Kornblihtt AR, Zorzopulos J (1993) Retroviral-like features in the monomer of the major satellite DNA from the South American rodents of the genus *Ctenomys*. DNA Seq 3:379–381

SanMiguel P, Bennetzen J (1998) Evidence that recent increase in maize genome size was caused by the massive amplification of intergene retrotransposons. Ann Bot 82:37–44

SanMiguel P, Tikhonov A, Jin YK, Motchoulskaia N, Zakharov D, Melake-Behan A, Springer PS, Edwards KJ, Lee M, Avramova Z, Bennetzen JL (1996) Nested retrotransposons in the intergenic regions of the maize genome. Science 273:765–769

Sanz-Alferez S, SanMiguel P, Jin Y-K, Springer PS, Bennetzen JL (2003) Structure and evolution of the Cinful retrotransposon family of maize. Genome 46:745–752

Siroky J, Lysak MA, Dolezel J, Kejnovsky E, Vyskot B (2001) Heterogeneity of rDNA distribution and genome size in *Silene* spp. Chromosome Res 9:387–393

Staden R (1996) The Staden sequence analysis package. Mol Biotechnol 5:233–241

Suoniemi A, Tanskanen J, Schulman AH (1998) Gypsy like retrotransposons are widespread in the plant kingdom. Plant J 13:699–705

Tek AL, Song J, Macas J, Jiang J (2005) Sobo, a recently amplified satellite repeat of potato, and its implications for the origin of tandemly repeated sequences. Genetics 170:1231–1238

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

Unfried K, Schiebel K, Hemleben V (1991) Subrepeats of rDNA intergenic spacer present as prominent independent satellite DNA in *Vigna radiata* but not in *Vigna angularis.* Gene 99:63–68

Vicient CM, Suoniemi A, Anamthawat-Jonsson K, Tanskanen J, Beharav A, Nevo E, Schulman AH (1999) Retrotransposon BARE-1 and its role in genome evolution in the genus *Hordeum*. Plant Cell 11:1769–1784

Zaki EA (2003) Plant retroviruses: structure, evolution and future application. Afr J Biotechnol 2:136–139

Zuker M (2003) Mfold web server for nucleic acid folding and hybridization prediction. Nucleic Acids Res 31:3406–3415