Plant Molecular Biology **34:** 255–264, 1997.

A *copia*-like retrotransposon Tgm*r* closely linked to the *Rps1*-k allele that confers race-specific resistance of soybean to *Phytophthora sojae*

Madan K. Bhattacharyya[1,*], Robert A. Gonzales[1], Michelle Kraft and Richard I. Buzzell[2]
[1]*Plant Biology Division, The Samuel Roberts Noble Foundation, P.O. Box 2180, Ardmore, OK 73402, USA (*author for correspondence);* [2]*Research Station, Agriculture Canada, Harrow, Ontario, Canada N6G 2V4*

Received 13 September 1996; accepted in revised form 21 February 1997

Key words: transactivation, insertional mutagenesis, RFLP-based screening, cosmid library, T-vector

Abstract

We have isolated and characterized Tgm*r*, a *copia*-like retrotransposon, linked tightly to the *Rps1*-k allele that confers race-specific resistance of soybean to the the fungal pathogen *Phytophthora sojae*. Southern analysis followed by PCR and sequence analyses, using primers based on sequences flanking the insertion site confirmed that the element was inserted in the neighboring region of *Rps1*-k but not in that of the other four *Rps1* alleles. This implies that Tgm*r* was transposed into the *Rps1*-k flanking site after the divergence of *Rps1* alleles. Southern analysis of a series of diverse soybean cultivars revealed a high level of polymorphism of Tgm*r*-related sequences. These results indicate that this low copy retroelement family could have been active in the soybean genome in the recent past. Tgm*r* contains long terminal repeats (LTR) and four non-overlapping open reading frames (ORF), presumably originating from mutations leading to stop codons of a single ORF. The conserved domains for gag, protease, integrase, reverse transcriptase and RNaseH are present in the internal portion of the element. However, the protease, reverse transcriptase and RNaseH of this element are non-functional due to the presence of several stop codons. Possible transactivation of Tgm*r* and application of this element in insertional mutagenesis for soybean are discussed.

Introduction

The retrotransposons are the largest class of mobile genetic elements. They typically encode reverse transcriptase, used in the synthesis of a new copy from an RNA intermediate produced by the host transcription machineries. During the transposition event, the original copy is left behind, thereby producing stable mutations. The retrotransposons are broadly classified into 2 groups based on the presence or absence of long terminal repeats (LTR): (1) LTR and (2) non-LTR retrotransposons. The LTR retrotransposons are again classified into 2 classes: (1) *copia* (*Drosophila*)- or Ty*1* (yeast)-like, and (2) *gypsy* (*Drosophila*)- or Ty*3* (yeast)-like elements. The *copia*- or Ty*1*-like elements are distinguished from *gypsy*- or Ty*3*-like elements based on the organization of reverse transcriptase and integrase. In the former class, integrase is located at the 5′ end of reverse transcriptase, whereas, in *gypsy*- or Ty*3*-like elements, integrase is located at the 3′ end of the RNaseH domain [9, 12]. Several *copia*-like plant retrotransposons have been isolated [4, 13, 20, 24, 28, 29, 39, 40, 42], and appear to be ubiquitous among plant species [10, 15, 43]. It has been suggested that *copia*-like elements are an ancient component of plant genomes, and may be involved in both gene duplication and regulation of gene expression [46]. Both non-LTR and *gypsy* or Ty3-like retrotransposons have also been documented in different plant species [25, 34, 37, 38, 48]. Recently, retrotransposon fragments representing *copia*-, *gypsy*,- and non-LTR- or LINE-like elements have been reported to account for $\geq$5% of mitochondrial genome of *Arabidopsis thaliana* [21].

Transposition of retroelements to a new target site causes stable integration of a new copy that can be

The nucleotide sequence data reported will appear in the EMBL, GenBank and DDBJ Nucleotide Sequence Databases under the accession numbers 496748.

scored as a dominant restriction fragment length polymorphism (RFLP). Lee *et al.* [24] have shown the application of such an element in genetic analyses of peas. In the present investigation while attempting to isolate RAPD markers that are linked to a soybean disease resistance gene *Rps1*-k [36], we isolated a marker OPRK15 which is part of a *copia*-like retrotransposon Tgm*r*. A survey of a diverse collection of soybean cultivars suggests that this retroelement may have been active in the soybean genome in the recent past. Possible transactivation of Tgm*r*, and application of this element in insertional mutagenesis for soybean, are discussed.

Materials and methods

Growing of seedlings and evaluation of an F_2 population for expression of the Rps1-*k gene*

Etiolated soybean cultivars or F_2 seedlings of the cross Elgin (*rps1*-k) × E420 (*Rps1*-k) were grown as described before [44]. Zoospores of *P. sojae* race 1 were used to inoculate etiolated hypocotyls. Hypocotyls were evaluated for symptom development 2 days after inoculation [44]. A total of 54 susceptible plants were selected for mapping molecular markers. In a separate experiment, F_2 seedlings of the cross E300 (*rps1*-k) × OX717 (*Rps1*-k) were grown in light and unifoliate leaves were inoculated for symptom development [1]. Leaves from 163 susceptible plants were harvested for DNA preparation.

RAPD analysis

RAPD analysis was carried out following the protocol of Williams *et al.* [47]. The decamer oligos were obtained from Operon Technologies, Inc. Alameda, CA. The OPRK15 marker is named after the oligo K15 with sequence CTCCTGCCAA. The OPRK15-specific band amplified from Williams 82, but not from Williams DNA templete was gel-purified and used as a probe in Southern blot analysis.

Nucleic acid isolation and Southern blot analysis

DNA was isolated from etiolated cotyledons or leaves following the procol of White and Kapper [45]. An 8 μg portion of DNA from individual samples was digested with individual restriction endonucleases, run on a 0.8% agarose gel and blotted onto a nylon membrane using 0.4 M NaOH, 1.5 M NaCl solution. The hybridization and washing of blots were carried out following standard protocols [27].

Mapping of molecular markers

Molecular markers were mapped using 54 susceptible F_2 seedlings of the cross Elgin (*rps1*-k) × E420 (*Rps1*-k). An additional 163 susceptible F_2 seedlings of the cross E300 (*rps1*-k) × OX717 (*Rps1*-k) were used in mapping OPRK15. The RFLP markers pA-71 and pA-280 were kindly provided by Dr R. C. Shoemaker, USDA-ARS, Ames, Iowa. The map distances and standard errors were calculated by the Map Manager program (email: mapmgr@mcbio.med.buffalo.edu).

Construction and screening of a cosmid library

High-molecular-weight DNA was isolated from the leaves of 2-week old plants as follows. Leaves were harvested and ground in an ice cold mortar and pestle with extraction buffer (0.1 M Tris-HCl pH 9.4, 0.1 M EDTA, 0.8 M KCl, 0.5 M sucrose, 1 mM phenylmethylsulfonyl fluoride (PMSF), 0.5% Triton X-100, 0.1% 2-mercaptoethanol, 40 mM spermidine, 10 mM spermine). The extract was filtered through a 100 μm nylon membrane. Cells in the filtrate were centrifuged down in a Sorvall RT6000 B centrifuge (DuPont, USA) at 2000 rpm, and the pellet was resuspended in extraction buffer without PMSF. Centrifugation and resuspension of cell pellets in buffer without PMSF were repeated 2 additional times. The cell pellet, after the final wash, was resuspended in an equal volume of extraction buffer without PMSF. This cell suspension was mixed with an equal volume of 2% agarose (low-melting-point agarose, Sea Plaque GTG, FMC bioproducts) in TE50 (10 mM Tris-HCl pH 9.4, 50 mM EDTA) at 45 °C. The mixture was placed in a mold (BioRad) and agarose plugs (76 μl) were obtained. The plugs were treated with proteinase K (1 mg/ml) in 0.5 M EDTA (pH 9.4) and 1% sarkosyl for 48 h at 55 °C with one change of enzyme solution. The plugs were treated with PMSF and then washed thoroughly with TE buffer before use. The high-molecular-weight (HMW) DNA was partially digested with *Mbo*I and size-fractionated in a sucrose gradient.

The size fraction carrying DNA fragments of ca. 50–75 kb was ligated to the Super Cos 1 vector (Stratagene), packaged, and around 240 000 cfu were obtained from ligation of 1 μg lambda arms to the soybean DNA inserts. Inserts of 20 randomly selected

clones revealed that each of the clones carried soybean DNA with an average insert size of 38 kb. The library was plated on 120 Petri plates (145 mm diameter) and grown overnight. The colonies from individual plates were stored in glycerol (18%) at −70 °C and used to prepare DNA. DNA from each cosmid pool was digested with *Hin*dIII and Southern blot filters were prepared. The blots were hybridized with OPRK15 to identify cosmid pools carrying this marker. Selected cosmid pools were screened by colony hybridization and individual cosmids carrying this marker were purified.

DNA sequencing and analysis

DNA was sequenced by the dideoxy sequencing method. A Taq DyeDeoxy Terminator Cycle Sequencing Kit (Applied Biosystems) was used according to the manufacturer's instruction. The PCR products were separated electrophoretically and the data were processed by an ABI373A automated DNA sequencer (Applied Biosystems). DNA and deduced amino acid sequences were analyzed using GCG programs.

PCR analysis of target sites

The DNA samples from soybean lines Mukden (*Rps1*-a), PI 84637 (*Rps1*-b), Lee 68 (*Rps1*-c) and PI 103.091 (*Rps1*-d) were PCR-amplified using two oligos based on nucleotide sequences flanking the insertion site (5′-end oligo GTAATCTCTTTATAGTATGCATG and 3′-end oligo GTTAGTAATAACCATGTTAC) at an annealing temperature of 50 °C for 1 min and extension at 72 °C for 5 min. The PCR products of less than 1 kb were purified from an agarose gel, and a second PCR step was carried out using a second pair of oligos (5′-end oligo GATTTGACATCATGTGAATTTGGAG and 3′-end oligo GTATGTTCAACGTGATCGACTTCACC) at an annealing temperature of 60 °C for 1 min and extension at 72 °C for 1 min. Ca. 215 bp long products from the sub-PCR reaction were cloned into a 'T' vector.

The 'T' vector used to clone the PCR fragments was prepared by digesting a modified pBluescriptII KS(+) plasmid with *Xcm*I to generate the 3′-end T overhangs. The plasmid had been modified by inserting one of two adaptors into the *Not*I/*Bam*HI and *Hin*dIII/*Xho*I sites of the plasmid, respectively. Adaptors were designed to have 2 *Xcm*I sites in opposite orientations to obtain T-overhangs at the 3′-end after *Xcm*I digestion. In addition to *Xcm*I sites, either an *Mlu*I or *Sfi*I site external to the *Xcm*I sites was included (Adaptor 1: GGCCGCTCTAGAACTAGTCCCACGCGTGCCATG/ GATCCATGGCACGCGTGGGACTAGTTCTAGAGC; Adaptor 2: AGCTTCCCATGCGAGGCCTGGATGGCCATCGATACCGTCGACC/ TCGAGGTCGACGGTAT CGATGGCCATCCAGGCCTCGCATGGGA). The β-galactosidase reading frame was maintained to allow blue/white selection of recombinant clones. The new plasmid vector is termed pRG51.

Results

Isolation and mapping of OPRK15

Near-isogenic lines (NILs) Williams and Williams 82 were used in identifying the RAPD marker OPRK15. This marker was obtained after screening 175 decamer-oligos in PCR reactions [47] (Operon Technologies, Alameda). Southern blot analysis of OPRK15 showed several common bands in both Williams and Williams 82, and two additional bands only in Williams 82 DNA digested with *Hin*dIII. This dominant marker, along with 2 previously mapped RFLP markers pA-71 and pA-280 [8] were mapped using 54 susceptible F_2 plants obtained from a cross between Elgin (*rps1*-k) and E420 (*Rps1*-k). OPRK-15 co-segregated with the *Rps1*-k gene. An additional 163 susceptible F_2 plants obtained from a different cross between E300 (*rps1*-k) and OX717 (*Rps1*-k) were evaluated for possible recombination between ORPK15 and *Rps1* loci. Only one recombination event was recorded between these 2 loci (Figure 1A).

OPRK15, pA-71, and pA-280 were also mapped using a series of NILs of different genetic backgrounds. These NILs differ at the *Rps1*-a, *Rps1*-c or *Rps1*-k alleles. A dominant resistance-specific OPRK15 band was observed in all lines carrying *Rps1*-k but not the *Rps1*-a or *Rps1*-c allele (Figure 1B). Analysis of Harosoy isolines carrying 5 different *Rps1* alleles also revealed that dominatnt OPRK15-specific bands were present only in lines carrying the *Rps1*-k allele (Figure 2). Sequence analysis revealed high identity of OPRK15 with the reverse transcriptase of Tnt*1*.

RFLP-based screening of cosmid pools for OPRK15

To characterize this sequence, a cosmid library carrying about 8 soybean genome equivalents of DNA was constructed in the Super Cos 1 vector (Stratagene)

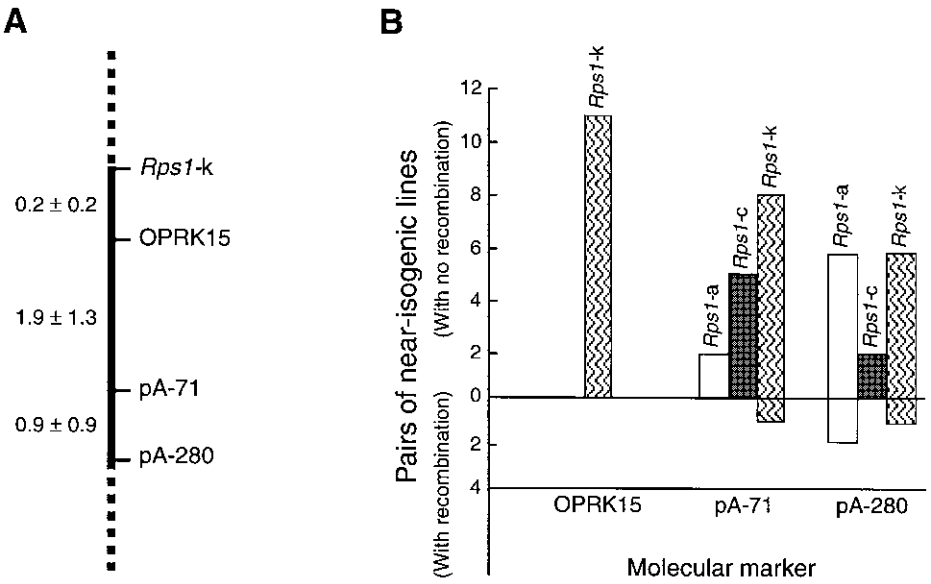

Figure 1. Mapping of molecular markers linked to *Rps1*-k. **A.** Genetic map position of OPRK15 in relation to the *Rps1*-k allele and RFLP markers. The map of pA-71 and pA-280 was based on the analysis of 54 susceptible F_2 plants of a cross between Elgin and E420. In addition to these 54 plants, 163 susceptible F_2 plants from the cross E300 × OX717 were used for mapping of OPRK15. For mapping OPRK15 and pA280 genomic DNA was digested with *Hin*dIII, while for pA-71 grenomic DNA was digested with *Taq*I. The map distances were calculated by using the Map Manager program. **B.** Mapping of molecular markers using near-isogenic lines. DNA from 14 pairs of NILs differing at *Rps1*-a, 8 pairs at *Rps1*-c, and 11 pairs at *Rps1*-k was digested with *Dra*I, *Taq*I and *Dra*I respectively for Southern blot analysis. Southern blots prepared from *Dra*I-digested DNA was hybridized to RFLP markers pA-280 and ORPK-15 individually, whereas blots carrying *Taq*I-digested DNA were hybridized to pA-71. Polymorphic band patterns between recurrent and donor parents are scored as allelic contrast, while similar band patterns between recurrent and donor lines are scored as non-allelic contrast. Pairs of lines showing allelic contrast are presented in the histogram. For example, only two out of 14 pairs of NILs differing at *Rps1*-a showed allelic contrast or polymorphism for pA-71. Dissimilar band patterns for a RFLP marker between the donor parent and the NIL carrying the introgressed region indicated a crossing over event between the *Rps1* and RFLP loci. For example, a single pair of NIL out of 11 pairs differing at *Rps1*-k showed recombination between *Rps1*-k and pA-71. Allelic contrast between donor and recurrent, but not between donor and introgressed NIL indicates the abscence of a recombination event between *Rps1* and RFLP loci. For example, 8 out of 11 pairs of NILs differing at the *Rps1*-k gene showed no recombination between *Rps1*-k and pA-71. The histogram for no recombination (upper part) indicates the pairs of lines that show allelic contrasts or polymorphisms between resistant and susceptible lines. The hisogram for recombination (lower part) indicates pairs of NILs that do not show allelic contrast among themselves but show allelic contrasts with their corresponding donor parents for the specific marker. OPRK15-specific band is absent in lines carrying either *Rps1*-a or *Rps1*-c.

and stored in 120 pools each carrying ca. 2000 cfu. DNA was isolated from each of these pools and digested with *Hin*dIII. Southern blot analysis of *Hin*dIII-digested pool DNA using OPRK15 identified 9 independent cosmid pools that carried OPRK15-specific sequences (Figure 3). Southern blot analysis of the cosmid 31 with 4 restriction endonucleases confirmed that this clone carries OPRK15. From this clone, OPRK15 specific *Hin*dIII, *Eco*RV, and *Dra*I fragments were subcloned into the plasmid vector Bluescript II (SK(−)) (Stratagene) for further characterization of this marker.

OPRK15 is a part of a copia*-like retrotransposon*

Sequencing of most of the ca. 6 kb *Dra*I fragment identified a 4965 bp *copia*-like retrotransposon element, Tgm*r*. This element has 2 almost identical long terminal repeats (5′-end LTR 260 bp; and 3′-end LTR 249 bp, 98% nucleotide identity between LTRs). LTRs carry an imperfect terminal inverted repeat sequence (5′-TcTTA.....TAAtA-3′) (Figure 4A). Tgm*r* also shares a 12 nucleotide (TGGTATCAGAGC) identity with bean $tRNA_i^{met}$ at the 5′ end [5] and carries a purine-rich sequence (GAGGGGGGG) adjacent to the 3′-end LTR, as in other plant retroelements [12]. There was a 5 bp (TAGGG) target site duplication upon insertion of this element. The element carries 4 non-overlapping ORFs, presumably resulting from at least 6 point mutations leading to 6 stop codons, if we assume that the point mutation leading to first stop codon between ORF2 and ORF3 caused the frameshift (see Figure 4B). As a result, the element does not carry the function-

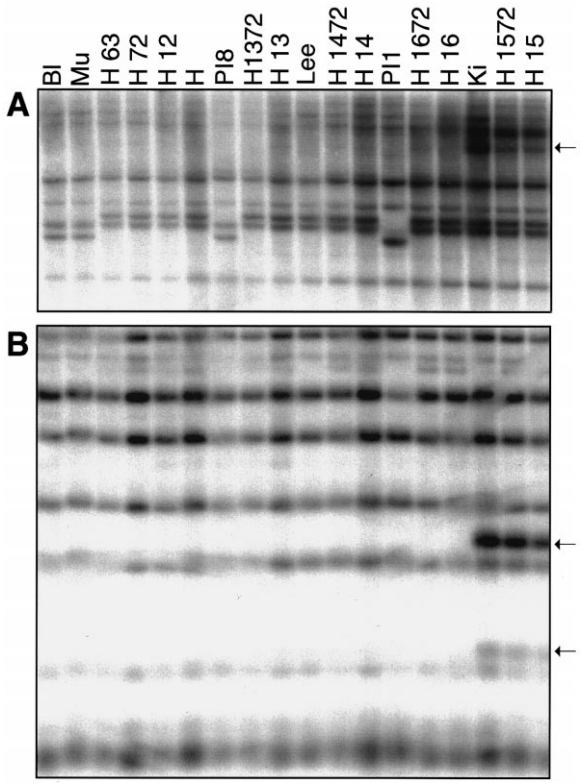

Figure 2. OPRK15 is linked to the *Rps1*-k allele but not to four other functional *Rps1* alleles. **A.** DNA from different Harosoy isolines carrying 5 different *Rps1* alleles and their respective donor parents was digested with *Bgl*II and Southern blots were hybridized to OPRK15. The Tgm*r*-specific fragment is shown with an arrow. This fragment was present in all the lines carrying the *Rps1*-k allele. Bl, Blackhawk (*Rps1*-a); Mu, Mukden (*Rps1*-a); H 63, (Harosoy 63, *Rps1*-a, *Rps7*); H 72, Harosoy (*rps1*-a, *Rps7*); H 12, HARO 12, (*Rps1*-a, *rps7*); H, HARO (1-7) (*rps1*-a, *rps7*); PI8, PI84637 (*Rps1*-b); H1372, HARO 1372 (*Rps1*-b, *Rps7*); H 13, HARO 13 (*Rps1*-b, *rps7*); Lee, Lee 68 (*Rps1*-c); H 1472, HARO 1472 (*Rps1*-c, *Rps7*); H 14, HARO 14 (*Rps1*-c, *rps7*); PI1, PI103.091 (*Rps1*-d); H 1672, HARO 1672 (*Rps1*-d, *Rps7*); H 16, HARO 16 (*Rps1*-d, *rps7*); Ki, Kingwa (*Rps1*-k); H 1572, HARO 1572 (*Rps1*-k, *Rps7*); H 15, HARO 15 (*Rps1*-k, *rps7*). **B.** DNA from different Harosoy isolines presented in A. was digested with *Hin*dIII and Southern blot analysis was carried out as in A. Tgm*r*-specific *Hin*dIII fragments are shown by arrows.

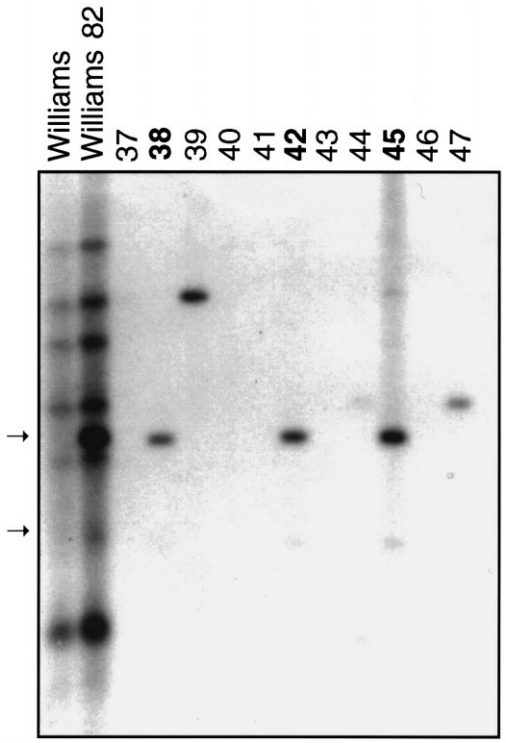

Figure 3. RFLP-based cosmid pool screening. DNA from each cosmid pool carrying ca. 2000 cfu was digested with *Hin*dIII and blotted along with the *Hin*dIII-digested DNA from NILs Williams and Williams 82. The OPRK15-specific bands are shown by arrows and pools carrying the OPRK15 sequences are shown in bold.

Table 1. Comparison of Tgm*r* with other *copia*-like elements.

Homology domain[2]	Amino acid identity with Tgm*r* (%)[1] Tnt*1-94*	Ta*1-3*	*copia*
GAG	26.3	29.2	23.7
INT	37.3	36.0	37.3
RT	38.3	34.0	36.1
RH	39.3	31.7	41.6

[1] Percent amino acid identity as measured by BESTFIT of the GCG package.
[2] GAG, gag protein; INT, integrase; RT, reverse transcriptase; RH, RNaseH.

al protease, reverse transcriptase and RNaseH found in other retrotransposons [9, 12] (Figure 4B). The element, however, carries all the essential internal domains that show significant identity to those of other *copia*-like retroelements *SIRE-1*, Tnt*1*, Ta*1* and *copia* [2, 13, 30, 42] (Table 1; Figure 4B; Figure 5). Based on similarity of general structure and identity of the amino acid sequences of Tgm*r* with that of other *copia*-like elements, we conclude that Tgm*r* is a non-autonomous *copia*-like retrotransposon.

Molecular analysis of donor parents carrying different Rps1 *alleles reveals that Tgm*r *is inserted only in the neighbouring region of* Rps1-*k*

Retrotransposons transpose through RNA intermediates and produce a DNA copy in a new chromosomal location by reverse transcriptase action. The original element is left behind and, therefore,

A.

```
Left Flank  TS     LTR                    Internal Domain        LTR                TS Right Flank

                   5'-End    3'-End                          5'-End    3'-End

Tgmr  TAAAAGGA TAGGG TcTTAGTC...TATTAAtA TGGTATCAGAGC...GAGGGGGGGC TcTTAGCA...TATTAAtA TAGGG AAAAGAGG

Target TAAAAGGA TAGGG                                                                  AAAAGAGG

tRNA_i^met              3'-End              ACCAUAGUCUCG
```

B.

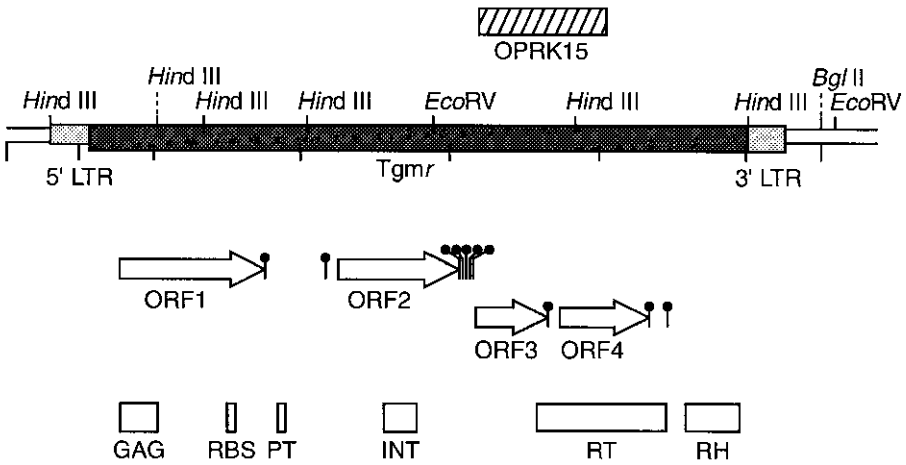

Figure 4. Organization of Tgm*r*. **A.** Salient features of the Tgm*r* element are presented. Flanking sites are based on the PCR-amplified product from line PI 103.091 (*Rps1*-d). TS, target site. Five bp imperfect terminal inverted repeats are underlined. The 5′ end of internal domain shows the 12 bp sequence identical to the 3′-end sequence of bean $tRNA_i^{met}$. The polypurine tract at the 3′ end of internal domain is in bold letters. **B.** Overall organization of Tgm*r* showing LTRs, 4 ORFs and internal domains. There are 10 mutations that resulted stop codons (•) between the gag and RNaseH and one frameshift mutation between ORF2 and ORF3. If we assume that the point mutation leading to first stop codon between ORF2 and ORF3 resulted in the frameshift, then 6 mutations for stop codons occurred between gag and RNaseH. RB, RNA binding motif; PT, protease; INT, integrase; RT, reverse transcriptase and RH, RNaseH. The positions of OPRK15 and different internal domains in Tgm*r* are shown.

copy number increases. However, due to rearrangement, internal domains may be deleted. For example, retrotransposon-like sequence Tms*1* of *Medicago sativa* carries 2 LTRs and a 163 bp internal domain [40]. To rule out the possibility that the observed lack of a Tgm*r* element linked to *Rps1* alleles other than *Rps1*-k could be a result of similar rearrangements in the flanking regions of other 4 *Rps1* alleles, equivalent flanking target regions of Tgm*r* from lines carrying different *Rps1* alleles were PCR-amplified in a sub-PCR reaction (see Materials and methods), and sequenced either directly or after cloning into a plasmid vector (see Materials and methods). The equivalent flanking target sequences from different lines confirmed the Southern blot data that Tgm*r* was inserted only in the flanking region of *Rps1*-k but not in that of other *Rps1* alleles (Figure 6).

Polymorphisms of Tgmr-related sequences among soybean cultivars indicate recent transposition of Tgmr-like elements in the soybean genome

We observed from analysis of several pairs of near-isogenic lines that Tgm*r* was inserted only in the flanking region of the *Rps1*-k allele (Figures 1B and 2). A diverse collection of 59 soybean cultivars, selected based on their geographical origin and different maturity groups, was analyzed for organization of Tgm*r* related sequences (Figure 7). We did not observe Tgm*r*-specific bands in any of the cultivars that do not carry the *Rps1*-k allele. Several polymorphic bands

Figure 5. Comparison of amino acid sequences across conserved domains of Tgm*r*, *SIRE-1*, Ta*1-3*, Tnt*1* and *copia*. Amino acid residues conserved among retrotransposons and retroviruses [13] are indicated by asterisks (*). Identical amino acid residues are shown in bold.

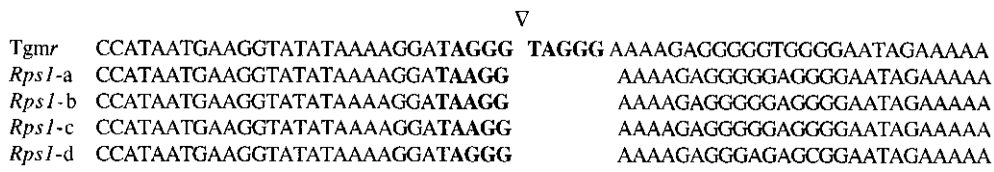

Figure 6. Flanking target sequences of Tgm*r* and sequences of equivalent regions from lines carrying other *Rps1* alleles. Tgm*r*, from Williams 82; *Rps1*-a, Mukden; *Rps1*-b, PI 84637; *Rps1*-c, Lee 68; *Rps1*-d, PI 103.091 (see Materials and methods for PCR and sequencing). ∇, position of Tgm*r*.

were observed among these cultivars. This indicates that Tgm*r*-like sequences may have been active in the recent past in the soybean genome.

Discussion

Genetic mapping using a large F_2 population and a series of near isogenic lines confirmed tight linkage of the RAPD marker ORPK15 with the *Rps1*-k allele. This marker hybridized to several additional genomic fragments in Southern blot analysis indicating the presence of a small family of OPRK15 related sequences in soybean. OPRK15 is a dominant marker and present only in lines carrying the *Rps1*-k allele. Absence of this dominant RFLP marker in soybean lines carrying *Rps1*-a, b, c and d functional alleles prompted us to further characterize this marker and the neighboring genomic region in order to isolate useful markers for analysis of different *Rps1* alleles. Furthermore, the OPRK15 sequence showed high identity to the reverse transcriptase of *Tnt*1 indicating that OPRK15 could be part of a retrotransposon. We were, therefore, interested in studying this element to explore the feasibility of using it in insertional mutagenesis of soybean.

Genomic cloning and sequencing revealed that ORPK15 is part of a soybean *copia*-like retrotransposon, Tgm*r*, that had transposed into the flanking region of the *Rps1*-k allele. Tgm*r* exhibits all of the characteristics of a *copia*-like element. The internal region carries 4 ORFs, presumably resulting from point mutations leading to 6 stop codons including 1

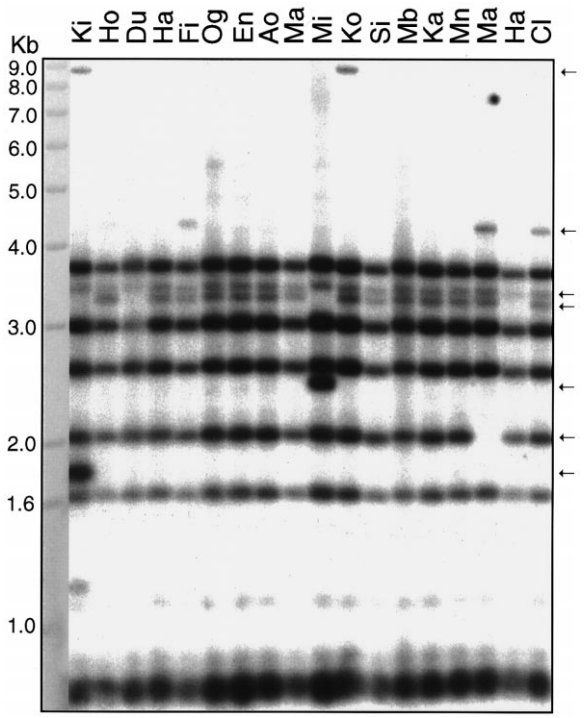

Figure 7. Organization of Tgm*r*-related sequences among soybean cultivars. Southern blots were prepared from DNA samples of 59 diverse soybean cultivars, digested with *Hin*dIII, and probed with OPRK15. This figure shows the polymorphic patterns of 18 random cultivars. Polymorphic bands are shown with arrows. Ki, Kingwa; Ho, Hodgson; Du, Dunfield; Ha, Habaro; Fi, Fiskeby V; Og, Ogemaw; En, Enrei; Ao, Aoda; Ma, Manchu; Mi, Midwest; Ko, Korean; Si, Sioux; Mb, Manitoba Brown; Ka, Kanum; Mn, Mandarin; Ma, Maple Aarrow; Ha, Harman; Cl, Cloud.

frameshift of a single ORF (Figure 4B). Analysis of the element, however, revealed all necessary domains normally present in *copia*-like elements (Figure 4B, Table 1). Additional retroelement-like sequences have been identified from soybean. The reverse transcriptase domain, isolated from soybean by Voytas *et al.* [43], showed 42.5% identity to the reverse transcriptase of Tgm*r*. Recently, a retrovirus-like element *SIRE-1* has been reported from soybean [2, 23]. Amino acid sequence comparison revealed no significant identity between *SIRE-1* and Tgm*r*.

Among the identified and characterized retroelements in plants, Tnt*1* and Tto*1* from tobacco, Bs*1* from maize, and Wis*2* from wheat were reported to be active [13, 14, 16, 19, 29]. Southern blot analysis indicates that Tgm*r* is member of a low copy number family of retrotransposons. Insertion of this element at the flanking region of *Rps1*-k allele but not that of other *Rps1* alleles implies that Tgm*r* was transposed into the *Rps1*-k flanking site after the divergence of *Rps1* alleles. Southern analysis revealed that Tgm*r*-related sequences have integrated into at least 6 additional locations polymorphic among soybean cultivars (Figure 7). These results indicate that the Tgm*r* family may have been active in the recent past. Considering the low copy number and an indication of probable transposition in recent time, a horizontal transfer of this family of retroelements can not be ruled out [22, 38].

Retrotransposons, isolated from several plant species, are represented by a large family and in most cases they are non-autonomous due to accumulation of mutations [17, 28]. Analysis of a large number of Tnt*1* RNAs present in tobacco protoplasts and roots revealed that the Tnt*1* RNA population is highly diverse. The major source of variability is considered to be the error-prone reverse transcription [6]. The error rate of reverse transcriptase from retroelements is not known. The error rate of purified reverse transcriptase from retroviruses ranges from 10^{-7} to 10^{-2} per nucleotide [33].

A large part of the mutations in a newly transposed retroelement can occur during the transcription of the progenitor copy [32]. It has been shown in yeast that the Ty*1* element transposes through an RNA intermediate [3]. Recently, transposition of Tnt*1* in *A. thaliana* has also been shown to occur through an RNA intermediate [26]. The *in vivo* error rate of RNA polymerase II is unknown. The *in vitro* error rate of wheat-germ RNA polymerase II ranges between 0.5×10^{-5} to 4×10^{-3} per nucleotide which is close to the *in vivo* and *in vitro* error rate of 10^{-5} to 10^{-4} per nucleotide for *Escherichia coli* RNA polymerase ([7], and references there in).

Molecular analysis revealed that Tgm*r* does not carry an active protease, reverse transcriptase or RNaseH. At least 6 point mutations leading to 6 stop codons occured in this 5 kb Tgm*r* element (Figure 4B). The mutation rate to generate 6 stop codons in Tgm*r* is around 50-fold higher than the mutation rate of 2.5×10^{-5} per nucleotide calculated for the Ty*1* element during a single cycle of transposition [11]. The actual rate of point mutation in Tgm*r*, in fact, has to be many-fold higher to generate 6 stop codons. It is unlikely that errors caused by both RNA polymerase II and reverse transcriptase together can account for all the accumulated point mutations in Tgm*r*, if its progenitor was an autonomous element. Therefore, it is possible that the progenitor element of Tgm*r* could be non-autonomous. If this is true, a transactivation mechanism may be suggested. Transactivation of inactive plant retroelements Bs*1* and *Athila* by reverse transcriptase

function from related elements have been suggested [18, 31]. It has also been shown that Ty*1*-encoded proteins can act in *trans* to activate a mini-Ty*1* element with no functional ORF [49].

Soybean, an agronomically important leguminous crop is rich in both protein and oil content. Poor transformation efficiency makes it difficult to apply T-DNA tagging or a map-based cloning approach in the isolation of genes with unknown gene products. In addition to the retrovirus-like element *SIRE-1*, the *Tgm* family of transposable elements have been characterized in soybean [35]. However, to date no active transposition event for the *Tgm* family, *SIRE-1*, or any other elements has been reported for this species [41]. Activation and application of a heterologous class II transposon, such as maize Ac-Ds elements, have not been demonstrated in soybean. Recently, it has been demonstrated that tobacco retrotransposon Tnt*1-94* was activated when overexpressed in *Arabidopsis thaliana* [26]. If Tgm*r* was *trans*-activated during its transposition, overexpression of Tgm*r* in soybean using a highly active promoter may activate Tgm*r* in the soybean genome. The results of such experiments will be invaluable in assesing the utility of this element in developing an insertional mutagenesis program for isolation of soybean genes with unknown functions. The mutation caused by Tgm*r* will be stable. However, since no revertents can be expected from such mutants, complementation analysis will be essential in the identification of Tgm*r*-tagged genes.

Acknowledgments

We thank Ann R. Harris and Angela D. Scott for assistance with oligonucleotide synthesis and DNA sequencing, Kirby D. Childs and Bonnie G. Espinosa for excellent technical assistance, Brandt G. Cassidy, Richard A. Dixon and Kenneth L. Korth for critical reading the manuscript. We thank Evan J. Burrows and Cuc Ly for photographs and illustrations and Allyson D. Wilkins for preparing the manuscript. The work was supported by the Samuel Roberts Noble Foundation.

References

1. Bhattacharyya MK, Ward EWB: Expression of gene-specific and age-related resistance and the accumulation of glyceollin in soybean leaves infected with *Phytophthora megasperma* f. sp. *glycinea*. Physiol Mol Plant Path 29: 105–113 (1986).
2. Bi Y-A, Laten HM: Sequence analysis of a cDNA containing the *gag* and *prot* regions of the soybean retrovirus-like element, *SIRE-1*. Plant Mol Biol 30: 1315–1319 (1996).
3. Boeke JD, Garfinkel DJ, Styles CA, Fink GR: Ty elements transpose through an RNA intermediate. Cell 40: 491–500 (1985).
4. Camirand A, St-Pierre B, Marineau C, Brisson N: Occurrence of a copia-like transposable element in one of the introns of the potato starch phosphorylase gene. Mol Gen Genet 224: 33–39 (1990).
5. Canaday J, Guillemaut P, Weil J-H: The nucleotide sequences of the initiator transfer RNAs from bean cytoplasm and chloroplasts. Nucl Acids Res 8: 999–1008 (1980).
6. Casacuberta JM, Vernhettes S, Grandbastien M-A: Sequence variability within the tobacco retrotransposon *Tnt*1 population. EMBO J 14: 2670–2678 (1995).
7. de Mercoyrol L, Corda Y, Job C, Job D: Accuracy of wheat-germ RNA polymerase II. General enzymatic properties and effect of template conformational transition from right-handed *B*-DNA to left-handed *Z*-DNA. Eur J Biochem 206: 49–58 (1992).
8. Diers BW, Mansur L, Imsande J, Shoemaker RC: Mapping *Phytophthora* resistance loci in soybean with restriction fragment length polymorphism markers. Crop Sci 32: 377–383 (1992).
9. Flavell AJ: Retroelements, reverse transcriptase and evolution. Comp Biochem Physiol 110B: 3–15 (1995).
10. Flavell AJ, Dunbar E, Anderson R, Pearce SR, Hartley R, Kumar A: *Ty1*-copia group retrotransposons are ubiquitous and heterogeneous in higher plants. Nucl Acids Res 20: 3639–3644 (1992).
11. Gabriel A, Willems M, Mules EH, Boeke JD: Replication infidelity during a single cylce of *Ty1* retrotransposition. Proc Natl Acad Sci USA 93: 7767–7771 (1996).
12. Grandbastien M-A: Retroelements in higher plants. Trends Genet 8: 103–108 (1992).
13. Grandbastien M-A, Spielmann A, Caboche M: Tnt1, a mobile retroviral-like transposable element of tobacco isolated by plant cell genetics. Nature 337: 376–380 (1989).
14. Hirochika H: Activation of tobacco retrotransposons during tissue culture. EMBO J 12: 2521–2528 (1993).
15. Hirochika H, Hirochika R: *Ty1-copia* group retrotransposons as ubiquitous components of plant genomes. Jpn J Genet 68: 35–46 (1993).
16. Hirochika H, Otsuki H: Extrachromosomal circular forms of the tobacco retrotransposon *Tto1*. Gene 165: 229–232 (1995).
17. Hu W, Das OP, Messing J: *Zeon-1*, a member of a new maize retrotransposon family. Mol Gen Genet 248: 471–480 (1995).
18. Jin Y-K, Bennetzen JL: Structure and coding properties of *Bs1*, a maize retrovirus-like transposon. Proc Natl Acad Sci USA 86: 6235–6239 (1989).
19. Johns MA, Babcock MS, Fuerstenberg SM, Fuerstenberg SI, Freeling M, Simpson RB: An unusually compact retrotransposon in maize. Plant Mol Biol 12: 633–642 (1989).
20. Kamm A, Doudrick RL, Heslop-Harrison JS, Schmidt T: 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 USA 93: 2708–2713 (1996).
21. Knoop V, Unseld M, Marienfeld J, Brandt P, Sünkel S, Ullrich H, Brennicke A: *copia-*, *gypsy-* and LINE-like retrotransposon fragments in the mitochondrial geneome of *Arabidopsis thaliana*. Genetics 142: 579–585 (1996).

22. Konieczny A, Voytas DF, Cummings MP, Ausubel FM: A superfamily of *Arabidopsis thaliana* retrotransposons. Genetics 127: 801–809 (1991).
23. Laten HM, Morris RO: *SIRE-1*, a long interspersed repetitive DNA element from soybean with weak sequence similarity to retrotransposons: initial characterization and partial sequence. Gene 134: 153–159 (1993).
24. Lee D, Ellis THN, Turner L, Hellens RP, Cleary WG: A *copia*-like element in *Pisum* demonstrates the uses of dispersed repeated sequences in genetic analysis. Plant Mol Biol 15: 707–722 (1990).
25. Leeton PRJ, Smyth DR: An abundant LINE-like element amplified in the genome of *Lilium speciosum*. Mol Gen Genet 237: 97–104 (1993).
26. Lucas H, Feuerbach F, Kunert K, Grandbastien M-A, Caboche M: RNA-mediated transposition of the tobacco retrotransposon Tnt1 in *Arabidopsis thaliana*. EMBO J 14: 2364–2373 (1995).
27. Maniatis T, Fritsch EF, Sambrook J: Molecular Cloning: A Laboratory Manual. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY (1982).
28. Manninen I, Schulman AH: *BARE-1*, a *copia*-like retroelement in barley (*Hordeum vulgare* L.). Plant Mol Biol 22: 829–846 (1993).
29. Moore G, Lucas H, Batty N, Flavell R: A family of retrotransposons and associated genomic variation in wheat. Genomics 10: 461–468 (1991).
30. Mount SM, Rubin GM: Complete nucleotide sequence of the *Drosophila* transposable element copia: Homology between copia and retroviral proteins. Mol Cell Biol 5: 1630–1638 (1985).
31. Pélissier T, Tutois S, Deragon JM, Tourmente S, Genestier S, Picard G: *Athila*, a new retroelement from *Arabidopsis thaliana*. Plant Mol Biol 29: 441–452 (1995).
32. Preston BD: Error-prone retrotransposition: rime of the ancient mutators. Proc Natl Acad Sci USA 93: 7427–7431 (1996).
33. Preston BD, Dougherty JP: Mechanisms of retroviral mutation. Trends Microbiol 4: 16–21 (1996).
34. Purugganan MD, Wessler SR: Molecular evolution of *magellan*, a maize Ty3/*gypsy*-like retrotransposon. Proc Natl Acad Sci USA 91: 11 674–11 678 (1994).
35. Rhodes PR, Vodkin LO: Organization of the *Tgm* family of transposable elements in soybean. Genetics 120: 597–604 (1988).
36. Schmitthenner AF: Phytophthora rot. In: Sinclair JB, Backman PA (eds) Compendium of Soybean Diseases, pp. 35–38. APS Press, St. Paul, MN (1989).
37. Schwarz-Sommer Z, Leclercq L, Göbel E, Saedler H: Cin4, an insert altering the structure of the *A1* gene in *Zea mays*, exhibits properties of nonviral retrotransposons. EMBO J 6: 3873–3880 (1987).
38. Smyth DR, Kalitsis P, Joseph JL, Sentry JW: Plant retrotransposon from *Lilium henryi* is related to *Ty3* of yeast and the gypsy group of *Drosophila*. Proc Natl Acad Sci USA 86: 5015–5019 (1989).
39. VanderWiel P, Voytas DF, Wendel JF: *Copia*-like retrotransposable element evolution in diploid and polyploid cotton (*Gossypium* L.). J Mol Evol 36: 429–447 (1993).
40. Vegh Z, Vincze E, Kadirov R, Tóth G, Kiss GB: The nucleotide sequence of a nodule-specific gene, Nms-25 of *Medicago sativa*: its primary evolution via exon-shuffling and retrotransposon-mediated DNA rearrangements. Plant Mol Biol 15: 295–306 (1990).
41. Vodkin LO: Plant transposable elements: potential application for gene tagging in soybean. In: Verma DPS, Shoemaker RC (eds) Soybean: Genetics, Molecular Biology and Biotechnology, pp. 69–89. CAB International Press, Wallingford, UK (1996).
42. Voytas DF, Ausubel FM: A *copia*-like transposable element family in *Arabidopsis thaliana*. Nature 336: 242–244 (1988).
43. Voytas DF, Cummings MP, Konieczny A, Ausubel FM, Rodermel SR: *copia*-like retrotransposons are ubiquitous among plants. Proc Natl Acad Sci USA 89: 7124–7128 (1992).
44. Ward EWB, Lazarovits G, Unwin CH, Buzzell RI: Hypocotyl reactions and glyceollin in soybeans inoculated with zoospores of *Phytophthora megasperma* var. *sojae*. Phytopathology 69: 951–955 (1979).
45. White JL, Kaper JM: A simple method for detection of viral satellite RNAs in small plant tissue samples. J Virol Meth 23: 83–94 (1989).
46. White SE, Habera LF, Wessler SR: Retrotransposons in the flanking regions of normal plant genes: A role for *copia*-like elements in the evolution of gene structure and expression. Proc Natl Acad Sci USA 91: 11 792–11 796 (1994).
47. Williams JGK, Kubelik AR, Livak KJ, Rafalski JA, Tingey SV: DNA polymorphisms amplified by arbitrary primers are useful as genetic markers. Nucl Acid Res 18: 6531–6535 (1990).
48. Wright DA, KE N, Smalle J, Hauge BM, Goodman HM, Voytas DF: Multiple non-LTR retrotransposons in the genome of *Arabidopsis thaliana*. Genetics 142: 569–578 (1996).
49. Xu H, Boeke JD: Localization of sequences required in *cis* for yeast Ty1 element transposition near the long terminal repeats: analysis of mini-Ty1 elements. Mol Cell Biol 10: 2695–2702 (1990).