View metadata, citation and similar papers at core.ac.uk

brought to you by CORE

provided by Okayama University Scientific Achievement Repository

Chemistry

Biochemistry fields

Okayama University *Year* 2004

Isolation of a transcriptionally active element of high copy number retrotransposons in sweetpotato genome

Makoto Tahara* Takahiro Aoki† Shinya Suzuka‡

Hiroki Yamashita** Masaru Tanaka††

Sachi Matsunaga‡‡ Shuhei Kokumai§

*Okayama University, tahara@cc.okayama-u.ac.jp

†JEOL Limited

‡SK Foods Company, Limited

**Okayama University

††National Agricultural Research Center

‡‡Kabaya Foods Corporation

§Okayama University

This paper is posted at eScholarship@OUDIR : Okayama University Digital Information Repository.

http://escholarship.lib.okayama-u.ac.jp/biochemistry/4

Running title: Isolation of a transcriptionally active element of high copy number retrotransposons in sweetpotato genome

Authors: Makoto TAHARA[a,1], Takahiro AOKI[b], Shinya SUZUKA[c], Hiroki YAMASHITA[a], Masaru TANAKA[d], Sachi MATSUNAGA[e] and Shuhei KOKUMAI[a]

[a] *Faculty of Agriculture, Okayama University, Okayama, Okayama, JAPAN 700-8530,*

[b] *JEOL, Akishima, Tokyo, JAPAN 196-8558,*

[c] *SK Foods, Kobe, Hyogo, JAPAN 651-2124,*

[d] *National Agricultural Research Center for Kyushu Okinawa Region, National Agricultural Research Organization, Miyakonojyo, Miyazaki, JAPAN 885-0091,*

[e] *Kabaya Foods, Mitsu, Okayama, JAPAN 709-2117*

[1]To whom correspondence should be addressed. E-mail: tahara@cc.okayama-u.ac.jp,
phone: 81-86-251-8312, fax: 81-86-251-8388.

Summary

Many plant retrotransposons have been characterized, but only three families (*Tnt1*, *Tto1* and *Tos17*) have been demonstrated to be transpositionally competent. We followed a novel approach that enabled us to identify an active element of the Ty1-*copia* retrotransposon family with estimated 400 copies in the sweetpotato genome. DNA sequences of Ty1-*copia* reverse transcriptase (RTase) from the sweetpotato genome were analyzed, and a group of retrotransposon copies probably formed by recent transposition events was further analyzed. 3'RACE on callus cDNA amplified transcripts containing long terminal repeats (LTR) of this group. The sequence-specific amplification polymorphism (S-SAP) patterns of the LTR sequence in the genomic DNA were compared between a normal plant and callus lines derived from it. A callus-specific S-SAP product was found into which the retrotransposon detected by the 3'RACE had been transposed apparently during cell culture. We conclude that our approach provides an effective way to identify active elements of retrotransposons with high copy numbers.

Keywords: retrotransposon, *Ipomoea batatas*, S-SAP, transposition

Introduction

Retrotransposons are transposable genetic elements that require the action of reverse transcriptase on an RNA intermediate to move through the genome (Kumar and Bennetzen 1999). They are divided into two groups, depending on the presence of long terminal repeats (LTR). LTR retrotransposons are further classified into the Ty1-*copia* and Ty3-*gypsy* families. LTR retrotransposons are major plant genome components. A number of them have been identified in various plant species, including *Bs1* in maize (Johns et al. 1985), *BARE-1* in barley (Manninen and Schulman 1993), *SIRE-1* in soybean (Laten et al. 1998), and *RIRE3* in rice (Kumekawa et al. 1999). These retrotransposons generally occur in large copy numbers in the host genomes and presumably have been multiplied in the course of the host plant evolution. However, most of these copies are defective in transpositional functions. To date, three families represented by *Tnt1* (Grandbastien et al. 1989), *Tto1* (Hirochika 1993), and *Tos17* (Hirochika et al. 1996) are the only retrotransposons for which transposition has been demonstrated.

The identification and characterization of an active member is crucial for the understanding of the regulation of the expression and transposition of a retrotransposon family (Takeda et al. 1999; Beguiristain et al. 2001). Active elements can be useful tools in plant genetics (Hirochika 2001). The three families of active plant retrotransposons have been identified as an inserted sequence in a mutated nitrate reductase structural gene (*Tnt1*, (Grandbastien et al. 1989)), or by RT-PCR based on conserved motifs of reverse transcriptase (RTase) of the Ty1-*copia* retrotransposons (*Tto1* (Hirochika 1993) and *Tos17* (Hirochika et al. 1996)). The RT-PCR products of the retrotransposon transcripts were screened in Southern blot analyses in which either a transposed copy or an increase in copy numbers that had occurred during tissue culture was determined using the RT-PCR products as probes (Hirochika 2001).

However, these methods have limitations when employed to isolate active retrotransposons from plant genomes. The first method is based on phenotypic alterations caused by a mutated gene and therefore is unspecific, as retrotransposon insertion is not the exclusive source of mutations. On the other hand, Southern blot analysis using RT-PCR products as probes may not be efficient if a large number of copies of an active element or if copies with cross-hybridizing sequences exist in the genome. In these cases, it will be difficult to identify the transposed copies by Southern blot analysis. In fact, the copy numbers of *Tto1* and *Tos17* have been reported to be less than 30 and 1 to 4, respectively. Another problem with this method is that the RT-PCR based on the conserved motifs amplifies transcripts not only from active but also from defective retrotransposons, and also read-through transcripts from cellular genes.

Among the three active plant LTR retrotransposons known, *Tto1* and *Tos17* transpose in response to tissue culture conditions. The present work aimed at the identification of tissue-culture-activated retrotransposons in sweetpotato (*Ipomoea batatas* (L.) Lam). As sweetpotato propagates vegetatively, meristem culture is a common practice to provide virus-free stocks as a seed source or for commercial vine production. Genetic differentiations in mericlones are often observed in morphological traits such as root skin color and shape (Shimonishi et al. 2001). These genetic alternations probably are related to the transposition of a retrotransposon that is induced during meristem culture. In this paper, we describe a novel approach that has led to the isolation of an active member of the Ty1-*copia* group of retrotransposons with a large copy number in the sweetpotato genome. We named this active retrotransposon *Rtsp-1* (Retrotransposon of sweetpotato).

Materials and Methods

*Sweetpotato Ty1-*copia *retrotransposon grouping based on the RTase sequence*

The RTase domain between the two conserved motifs of the sweetpotato Ty1-*copia* retrotransposon was previously cloned and sequenced (Tanaka et al. 2001). Together with the RTase sequences of sweetpotato found in the DNA Data Bank of Japan (DDBJ, http://www.ddbj.nig.ac.jp/), 50 sequences in total were analysed. After aligning the sequence data with the Clustal W Multiple Sequence Alignment Program (Thompson et al. 1994), 1,000 data sets were produced by bootstrap resampling (Felsenstein 1985) and a genetic distance matrix was calculated for each data set based on Kimura's 2-parameter method where the transition/transversion parameter was set to 2.0 (Kimura 1980). A consensus neighbor-joining tree of the bootstrap data sets was estimated (Saitou and Nei, 1987). PHYLIP was used for all phylogenetic analyses (Felsenstein 1993).

Mericlone culture and callus development

Shoot tips (2 cm) were harvested from greenhouse-grown plants of the sweetpotato cultivar Koukei 14, and were surface-sterilized with ethanol and sodium hypochlorite. Shoot apices (0.3 mm) were isolated from the shoots and cultured on MS medium (Murashige and Skoog 1962) supplemented with 3% sucrose, 0.2 mg/L 1-naphthaleneacetic acid, and 2 mg/L 6-benzyladenine, solidified by 0.8% agar, under 10 hours light per day at 25°C. Shoots, leaves and roots developed normally from an apical tip, and callus tissue formed at its basal part. A few apices showed abnormal leaf and root formation. The calli formed at the apex showing normal (KB-1) or abnormal (KB-4) differentiation were harvested and cultured separately as described above. They were maintained by transferring them onto fresh medium every two months.

Nucleic acid extraction

Genomic DNA was extracted from young leaves of Koukei 14 plants, and from KB-1 and KB-4 callus tissue by the cethytrimethylammonium bromide (CTAB) method described by Hayakawa (1997). Total cellular RNA from KB-4 callus tissue was isolated using a modified guanidine thiocyanate method (Totally RNA; Ambion, Austin, TX) according to the manufacturer's recommendations. The RNA extract was subjected to PCR using beta-actin primers (5'-CGCGGATCCGARAARATGACNCARATHATGTT-3' and 5'-AAACTGCAGATRTCNACRTC RCAYTTCAT-3'), which were set to amplify the fragment between two exons of the beta-actin gene, to ensure the absence of detectable DNA contamination.

3'RACE

3' RACE analysis (Frohman et al. 1988) was carried out on cDNA prepared from total RNA of KB-4 callus tissue. ReverTra Ace (Toyobo, Osaka) was used for reverse-transcription with the oligo(dT) primer (5'- GACCACGCGTATCGATGTCGACTTTTTTTTTTTTTTTV-3') following the manufacturer's directions. The first PCR amplification of the cDNA was performed with the group 1 (Gr1) specific primer (Gr1, 5'-GAAGGATTTGAAGACAAAGAGAATC-3' as in Fig. 1b) and an adapter primer (5'-CCACGCGTATCGATGTCGACTTTTTTTT-3'). The second amplification was performed with the nested Gr1 downstream (5'-AGGTTGAACAAATCTCTRTACGGTT-3' as in Fig. 1b) or the RNaseH (5'-TGGATTCTGACTATGCAGGGG-3' as in Fig. 2c) primer together with the adapter primer mentioned above.

PCR

When short (i.e., less than 1,500 bp) PCR products were expected, *Taq* DNA polymerase (Sigma-Aldrich St. Louis, MO) was used in a cycling profile of 94°C for 3 min, 35 cycles of 94°C for 1

min, 55°C for 1 min and 72°C for 2 min, and a final extension at 72°C for 7 min. The annealing temperature was adjusted to meet the requirements of the primers when necessary. For long expected PCR products, *ExTaq* (TaKaRa, Tokyo) was used following the manufacturer's direction.

Dot blotting analysis

Dot blotting hybridization was employed to determine the copy number of the *Rtsp-1* retrotransposon in the sweetpotato genome. A DNA fragment consisting of an RTase domain insert of 600 bp and a portion of a plasmid multiple cloning site of 19 bp was obtained as an RT-PCR plasmid clone digest by *EcoR*V and *BamH*I (Fig. 2c). The fragment was purified by phenol/chloroform extraction and separated from the vector by agarose gel electrophoresis before it was labeled with digoxigenin by random priming using the DIG High Prime DNA Labeling Kit (Roche Applied Science, Indianapolis, IN). Serial dilutions of the RT-PCR plasmid and genomic DNA were spotted onto Hybond N^+ membranes (Amersham Biosciences, Piscataway, NJ) with three replications. The plasmid standard spots contained 0.001, 0.01, 0.1, 1 and 10 ng of DNA, whereas genomic DNA dots contained 40, 200, and 1000 ng. DNA was quantified on agarose gels stained with ethidium bromide. Hybridization was performed according to the manual of the DIG High Prime DNA Labeling Kit, with a final stringency wash of 0.1 x SSC and 0.1% SDS at 68° C for 15 min. Hybridization signals were quantified with a Fluor-S MAX MultiImage (BioRad Laboratories Japan, Tokyo). The dependence of the hybridization signals on the DNA quantity of plasmid standards was nearly perfectly linear in the range of 0.001 to 1 ng, and the hybridization signals for the genomic dots fell within this range. The relative quantity of *Rtsp-1* retrotransposon DNA present in the sweetpotato genome (1,597 Mbp/1C, 2n=6x=90, Arumuganathan and Earle 1991) was determined, and the the copy numbers of the *Rtsp-1* retrotransposon was estimated.

S-SAP

Total DNA (5 µg) was digested with 2 U of *Mse*I for 12 h in NE buffer 2 (10 mM TRIS-hydro chloride pH 7.9, 10 mM magnesium chloride, 50 mM sodium chloride, 1 mM dithiothreitol, 100 ng/µL BSA) in a total volume of 50 µL. Digested DNA was purified using a Phenol:Chloroform:Isoamyl Alcohol 25:24:1 mixture and was suspended in 50 µL TE buffer (10 mM TRIS-hydro chloride pH 7.5, 1 mM EDTA). A volume of 5 µL of this DNA solution was ligated for 2 h at 16°C with 15 pmol MseI adapter (5'-GACGATGAGTCCTGAG-3' annealed with 5'-TACTCAGGACTCAT-3', Waugh et al. 1997) using 1 U of T4 DNA ligase (New England Biolabs, Beverly, MA) in ligation buffer (50 mM TRIS-hydrochloride pH 7.5, 10 mM magnesium chloride, 10 mM dithiothreitol, 1 mM ATP, 25 ng/µL BSA) in a total volume of 20 µL. In order to selectively amplify and label the fragments containing the 5' end of the *Rtsp-1* LTR and the adjacent host sequence, two rounds of PCR using nested primers were run. The first PCR was carried out in 30 µL containing 5 µL of the ligation mixture; the primer combination consisted of the PBS_Comp primer containing a complementary sequence of the *Rtsp-1* primer binding site (PBS) (5'-CAGAAGGCTCTGATACCAATTGTTGCGC-3', Fig. 2b), and the M primer, homologous to the adapter sequence (5'-GACGATGAGTCCTGAGTAA-3'). The second PCR was performed in 10 µL containing 0.4 µL of 40-fold diluted first PCR products using the nested LTR primer, Ext_R (5'-CCACTCTCTAACTAACAAGGA-3', Fig. 2b) with Texas Red labeling at the 5' end, and the M plus A (adenine) selective primer, with adenine added at the 3' end of the M primer sequence. Both PCR mixtures contained 0.08 U/µL *Taq* DNA polymerase (Sigma), 10 mM TRIS-hydro chloride pH 8.3, 50 mM potassium chloride, 2.5 mM magnesium chloride, 200 µM dNTPs and 40 nM of each primer. The cycling profile was 94°C for 1 min, followed by 30 cycles of 94°C for 1 min, 63°C for 1 min, and 72°C for 2 min, and finally 72°C for 5 min. After PCR, samples were denatured at 95°C for 2 min

and immediately placed on ice. Gel loading buffer (0.5 µL; 94% formamide, 10 mM EDTA, 0.5 mg/ml bromophenol blue) was added to each denatured sample (2.0 µL), which was then loaded onto a 6% denaturing polyacrylamide gel. Electrophoresis was performed and the fluorograph image was created using a Hitachi SQ5500 sequencer (Hitachi, Tokyo).

Cloning, colony selection, and DNA sequencing

RT-PCR products were resolved on agarose gels; the bands were excised, purified with the QIAquick gel extraction kit (Qiagen, Chatsworth, CA), and cloned using the TA cloning kit (Invitrogen, Carlsbad, CA). In the case of the S-SAP products, the second S-SAP PCR was performed using a non-labeled Ext_R primer instead of the labeled primer to ensure ligation efficiency of the products into the TA cloning vector. The PCR products of the KB-4 callus were electrophoresed in an agarose gel. The products with sizes matching that of the target product were excised and cloned as described for the RT-PCR products. PCR reaction mixture was prepared as described for the second S-SAP except that the template DNA was not included. Transformed *Escherichia coli* colonies were individually picked from the culture plate, transferred to a fresh plate and suspended in 10 µL of the PCR mixture using a sterile toothpick. An initial denaturing step was included with the first PCR cycle to disrupt the *E. coli* membrane and inactivate its nucleases. The PCR amplified and labeled the S-SAP product that had been inserted into the bacteria. The amplified insertion products were electrophoresed together with the S-SAP products of the KB-4 callus as a reference using a Hitachi SQ5500 sequencer. Based on fluorograph images, colonies that appeared to contain plasmids of the same size as the callus-specific S-SAP product were selected for amplification. Plasmid DNA was isolated and sequenced on an ABI310 automated sequencer using the ABI PRISM BigDye Terminator Cycle Sequencing Ready Reaction kit (Applied Biosystems, Foster City, CA).

Results and discussion

Grouping of RTase sequences of Ty1-copia *retrotransposons in sweetpotato genome*

A total of 50 DNA sequences corresponding to the inside of the two conserved amino acid motifs, DVKTAFL (Voytas et al. 1992) and YVDDM (Hirochika et al. 1992), of the sweetpotato Ty1-*copia* retrotransposon RTase domain were analyzed and a neighbor-joining tree was constructed (Fig. 1a). There were at least four groups (Gr1-4; Fig. 1a) of sequences with bootstrapping probabilities of over 99%. The DNA sequences of recently transposed copies should be nearly identical even if the error-prone nature of reverse-transposition is taken into account. Retrotransposition proceeds by transcription from DNA to RNA, reverse transcription from RNA to DNA, and, finally, the synthesis of double-stranded DNA prior to integration into the host genome. Cellular RNA polymerases catalyze the first step, whereas reverse transcriptases do the rest. Reverse transcriptases as well as RNA polymerases lack exonucleolytic proofreading activity and, therefore, a high error rate must be expected in this type of replication (Preston 1996). On the other hand, the sequences of copies transposed in the distant past are likely to have been differentiated by non-directional and independent mutations. The degree of differentiation within a family of copies therefore should be proportional to the age of the individual transposition events. We concluded that group Gr1 was likely to have been formed by recent transposition events because of its extraordinarily dense clustering (Fig. 1a). Therefore, this group was considered the most likely one to have retained transpositional activity. Consensus DNA sequences were determined for each group and 3' RACE primers were designed to amplify the transcripts of Gr1 but not those of the other groups (Fig. 1b).

3'RACE products of Rtsp-1 *in callus tissue*

Two of the 31 shoot apices of the sweetpotato cultivar Koukei 14 that were cultured on MS medium showed abnormal leaf and root formation. The calli formed at the basal part of the apex showing normal (KB-1) or abnormal (KB-4) development were harvested and cultured. After six to eight passages of the culture, calli were harvested and nucleic acids were extracted.

3‘ RACE was performed on cDNA prepared from total RNA of KB-4 callus tissue using the oligo(dT) primer. The Gr1-specific primer and an adapter primer were used for the first 3’ RACE amplification, followed by the semi-nested PCR with the Gr1 downstream primer and the same adapter primer. The semi-nested PCR generated a 3’ RACE product of 919 bp (3’ RACE-1 in Fig. 2c). The sequence of the product showed significant homology to the RTase and RNaseH domains of *Tto1* and *Tnt1*, but the product was polyadenylated in the RNaseH domain (Fig. 2c). An RNaseH-specific primer was then designed and a semi-nested PCR was run using the first 3’RACE reaction mixture. This PCR amplified a product of 914 bp (3’ RACE-2 in Fig. 2c) which contained a sequence of 205 bp overlapping with the first 3’RACE product as well as the conserved end sequence of a polypurine tract (PPT) (5’-GGGAG-3’) followed by 5’-TG-3’ with two intervening nucleotides. The product was polyadenylated at the position 353 nucleotides downstream of the PPT end. This span should correspond to the U3 (unique at 3’) and R (repeated terminus of a transcript) regions of LTR since the U3 region starts with 5’-TG-3’, and the putative TATA motif at position 160-165 is followed by the presumable cap signal for transcription initiation.

A pair of primers was designed to amplify the fragment between the 5’ LTR (LTR primer: 5’-TCCTTGTTAGTTAGAGAGTGGTTGTAAC-3’) and the RTase (Gr1 RTase anti-sense primer: 5’-CAAAGTTATCTTTTCCAAACCTCTTG-3’) of the transcribed retrotransposon. After cloning the

PCR product generated with these primers from the genomic DNA of young leaves of Koukei 14, the sequence of the U5 (unique at 5') region of LTR and PBS was determined; 13 nucleotides of the PBS were found to be complementary to a sequence at the 3' end of the initiator methionine tRNAs of soybean (DDBJ/EMBL/GenBank accession number (ACC#): M17129), wheat (*Triticum aestivum*, ACC#: V01383), *Phaseolus vulgaris* (ACC#: K00318) and yellow lupin (*Lupinus luteus*, ACC#: X06458).

Copy numbers of the Rtsp-1

The relative DNA quantity of *Rtsp-1* in the sweetpotato genome was estimated to be 0.0205% (S.D. 0.0021%) by dot blotting hybridization. Assuming that the haploid genome size was 1,597 Mbp (Arumuganathan and Earle, 1991), 0.0205% of genomic DNA in hexaploid sweetpotato genome would be equivalent to 396 copies of *Rtsp-1* with the size of 4,968bp. Insertion sites of different retrotransposon transposition events are independent, implying that repeated insertions at identical sites on different homologous chromosomes are unlikely. In self-pollinating species, each insertion segregates after a series of self-pollinations; some offspring plants would have the identical insertions on homologous chromosomes, while the others would have none at the same chromosome site. Since the most retrotransposons have remained inactive over many generations of host plants, in the case of self-pollinating diploid species, if the insertion is observed at a particular chromosome site, the inserted copy most likely has been duplicated at the identical site of a homologous chromosome. Sweetpotato shows complete self-incompatibility. Any individual plant of sweetpotato will probably have identical insertion sites on homologous chromosomes only if the parental plants share a common ancestor in which the insertion event in question took place. However, the probability of such a combination of

identical insertion sites by sexual recombination is reduced in sweetpotato, which usually propagates vegetatively. Therefore, most of the about 400 copies of *Rtsp-1* can be expected to be spread over the sweetpotato genome, as the occurrence of homologous copies at identical insertion sites is unlikely.

Isolation of a transposed copy of the Rtsp-1 *retrotransposon*

The S-SAP procedure based on the *Rtsp-1* retrotransposon LTR amplified numerous products from Koukei 14 and callus genomic DNA (Fig. 3). The A (adenine) added selective M primer was used in the PCR for the S-SAP shown in Fig. 5b. We tested three other selective primers, M plus T (thymine), plus C (cytosine), or plus G (guanine), with the same set of samples. The banding patterns of the products largely differed between the primers used (fluorographs not shown). If a selective primer is used in the second S-SAP PCR, the S-SAP fragments in which the nucleotides flanking the *Mse*I restriction site match the selective nucleotide of the primer will be preferentially amplified. The results obtained agreed with the dot blotting estimates of the *Rtsp-1* copy numbers in the sweetpotato genome. The S-SAP banding patterns were almost identical between Koukei 14 plants and the two callus lines derived from this cultivar, but a few callus-specific S-SAP products also were found (Fig. 3). The S-SAP products of the KB-4 callus were cloned, and a colony transformed with a plasmid containing a 624 bp product derived from KB-4 callus was selected using the S-SAP procedure for plasmid insertion size determination on a sequence gel. The product contained the 5' terminal LTR sequence at one end and a flanking host sequence with the *Mse*I restriction site at the other ("S-SAP 624 bp" in Fig. 5b). The flanking genome sequence of this S-SAP product was homologous to the *CLAVATA1* (*CLV1*)-like receptor kinase of soybean (Yamamoto et al. 2000). PCR using primers homologous to sequences of *CLV1* (CL1 primer, 5'-ATACATGCCTAATGGGAGCTTGTCG-3') and LTR (U3 primer,

5’-GGGACCAAATGATAATTGTAATTGTCAACACAATTT-3’) with genomic DNA as the template produced the fragment expected from the S-SAP sequence (Fig. 5b) for KB-4, but not for KB-1 or Koukei 14 (Fig. 4). The same PCR did not generate fragments specific for any other sweetpotato cultivars or callus lines tested (data not shown). This confirms the the 624 bp S-SAP product is unique for KB-4 callus and suggests that this S-SAP site most likely was generated by transpositional insertion of the *Rtsp-1* retrotransposon during KB-4 tissue culture.

The S-SAP technique has been exploited to identify PCR fragments derived from newly transposed copies of *Tnt1* in regenerated plants (Melayah, et al, 2001). The copy number of *Tnt1* has been estimated as several hundreds (Grandbastien et al, 1989), and the S-SAP was demonstrated to be a powerful tool to detect transposition in highly repetitive transposon families. Recently, the S-SAP technique was successfully used to identify the first active DNA transposon from rice, *Pong*, and the first active miniature inverted-repeat transposable element (MITE), *mPing* (Jiang, et al, 2003). Moreover, the S-SAP procedure has another advantage. The S-SAP fragments derived from newly transposed copies can be cloned without extensive laboratory procedures as the same S-SAP procedure is applied for plasmid insertion size determination.

Data base searches indicated that the *CLV1*-homologous part of the S-SAP product was over 85% identical to CLV1-sequences of *Glycine max* (ACC#: AF197946), *Pisum sativum* (ACC#: AJ495759), and *Lotus japonicus* (ACC#: AJ495844). Since the *CLV1* homolog was interrupted by the *Rtsp-1* insertion, the sequence flanking the 3’ LTR terminal was deduced from these homologous *CLV1*-like genes, as this sequence was identical over a stretch of 35 nucleotides in these genes except for a single position. PCR with KB-4 genomic DNA using the RNaseH primer and a primer deduced from the

putative flanking sequence (5'-CAGCAATRGAGGACATGGACTGAGA-3') produced a single fragment, from which the actual flanking sequence was determined.

PCR using genomic DNA and primers with the sequences flanking the transposed *Rtsp-1*, CL1 and CL1_anti (5'- AACTCCAACATTGTTGAAGGATTTAGCCA-3'), generated at least two products of 241 bp and 4,791 bp from Koukei 14 and KB-1, but of 241 bp and 5,214 bp from KB-4 (Fig. 5a). The 241 bp fragment was a *CLV1* homolog without retrotransposon insertion, whereas the 4,791 bp product contained parts of two tandem aligned *CLV1*-like genes (not shown, registered in the DDBJ/EMBL/GenBank as AB162660). More specifically, the fragment consisted of the terminal part of the exon1, intron, and exon2 of the first *CLV1* copy, an intergenic spacer, and a big part of exon1 of the second *CLV1* copy. The sequence corresponding to exon1 of the second *CLV1* copy was 2,553 bp long without unexpected stop codons.

The 5,214 bp fragment contained the retrotransposon sequence inserted into exon1 of the second *CLV1* copy. The size of the inserted sequence was 4,968 bp flanked by 5 bp direct repeats (5'-TTTGG-3') as in Fig. 5b. This type of duplication is typically formed during retrotransposon integration into the host genome because the cleavage sites of the host DNA strands are staggered. After initial ligation between the 3' ends of the retrotransposon DNA and the 5' ends of the host DNA, short single-stranded sequences are left at both sides of the integrated retrotransposon DNA, which are later completed. This target site duplication provides additional and conclusive evidence for the *Rtsp-1* transposition in the callus genome; the junction between the *Rtsp-1* and the *CLV1* homolog observed in the KB-4 genome had been formed by transpositional insertion of *Rtsp-1*, but not by DNA recombination or chromosomal duplication. In addition, the flanking genome sequences (210 bp + 36 bp) perfectly matched that of exon1 of the second *CLV1* homolog in the 4,791 bp product (Fig. 5b).

The transposed retrotransposon had a sequence highly identical with the 3'RACE product (1,532/1,552 base pair match, Fig. 2c). Therefore, the inserted sequence is a member of a retrotransposon family which has maintained transcriptional activity in the sweetpotato genome. This result underlines the success of our strategy: running the 3'RACE with sequences of retrotransposons selected on the basis of the putative transposition history facilitates identification of active elements in high copy number retrotransposons in plants.

Although *Rtsp-1* was inserted into a *CLV1* homolog, the 241 bp fragment was detected in KB-4. This suggests that *CLV1* homologs other than the one carrying the *Rtsp-1* insertion were present in the homologous chromosomes. On the other hand, the 4,791 bp fragment was absent from KB-4 suggesting that the tandem-aligned *CLV1* fragment occurred in only one of the homologous six chromosomes, where the *Rtsp-1* insertion increased its size.

The transposed retrotransposon copy is considered to be full length and is registered as *Rtsp-1* in DDBJ/EMBL/GenBank (ACC#: AB162659). The copy had LTRs of 472 bp which differed only in one single nucleotide. The highly conserved TATA sequence started at position 160 of the LTR (Fig. 2b). The coding region between the two LTRs encoded 1,298 amino acids for *gag* and *pol* proteins in a single translational reading frame. The encoded protein shows high homologies with *Tnt1* (43% amino acid identities, bit score: 1083) and *Tto1* (41% identities, bit score: 1038), and has the typical structural feature of the Ty1-*copia* element, where the order of functional domains in *pol* is protease, integrase, RTase, and RNaseH (Fig. 2a).

The 3'RACE procedure with RTase primers amplified a polyadenylated fragment from the RNaseH domain. Since this fragment had a sequence that was nearly identical with that of the transposed copy (Fig. 2c), its precursor RNA appeared to be a transcript from the same retrotransposon as the full-length

RNA. According to the transposed copy, the sequence extending in 3' direction from the polyadenylated site was rich in thymine (Fig. 2c). This excludes the possibility that the fragment was an artifact caused by non-specific binding of the primers used in the 3'RACE. However, no polyadenylation sequence signals (Li and Hunt 1997) including near-upstream elements (NUE) and cleavage sites (CS) were identified. The polyadenylation process and possible biological role of this product therefore remain unclear.

A copy of the *Rtsp-1* retrotransposon was transposed into the exon1 of the *CLV1* homolog in KB-4 sweetpotato callus during tissue culture. Most of the corresponding exon1 (2,553 bp) was sequenced, but no unexpected stop codon was found. This suggests that this *CLV1* homolog is a functional gene. In *Arabidopsis thaliana*, three CLAVATA genes, *CLV1*, *2*, and *3*, regulate meristem development. *CLV1* mutant plants accumulate undifferentiated cells at the shoot and flower meristems resulting in stem overgrowth and the production of extra flowers (Clark et al, 1997). *Rtsp-1* insertion into the *CLV1* homolog might have caused a loss-of-function mutation, which has led to the abnormal development of the KB-4 plants.

Tissue culture and retrotransposition

In addition to the three transpositionally active retrotransposon families mentioned above, transcription has been reported for LTR retrotransposons in various plant species (Grandbastien, 1998). Using LTR and reporter gene constructs in a transgenic system, transcriptional activity of LTR sequences was further demonstrated for some of these retrotransposons including *BARE-1* in barley (Suoniemi, et al., 1996), *Panzee* in pigeonpea (Lall, et al., 2002) and *MCIRE* in alfalfa (Ivashuta, et al., 2002). However,

transcription of LTR retrotransposons is essential but not sufficient for retrotransposition. New insertions into the genome by transposition have actually not yet been identified in these transcriptionally active families. There are five major control points of the transposition cycle of LTR retrotransposon: transcription from DNA to RNA, translation and control of the *gag*/*pol* ratio, RNA packaging and virus-like particle assembly, reverse transcription to cDNA, and integration into host genomes (Grandbastien, 1998). Restriction at any of these steps may hinder transposition. Transposition of *Tnt1*, *Tto1* and *Tos17* is primarily regulated at the transcription step; transcription occurs in response to wounding or pathogen attack (*Tnt1* and *Tto1*), during tissue or cell culture (*Tto1* and *Tos17*), or in roots (*Tnt1*). As transcription of *Tnt1* and *Tto1* has been detected by placing a plant under stress conditions, it seems possible to detect transposition in naturally growing plants as a stress response i.e., after a systemic virus infection. However, to our knowledge, the transposition is reported only in callus tissue or the plants derived from tissue or cell culture. The transposition of the *Rtsp-1* is also found in callus tissue. Although this may simply reflect technical difficulty in detection of the transposition occurred in stress-treated plants, other currently not identified restrictions on retrotransposition might be temporally lifted in the cell during tissue or cell culture, when major modifications of cell metabolism and gene expression take place (Grandbastien, 1998).

Acknowledgements

This work was supported by a grant-in-aid for scientific research (C 12660007) from the Ministry of Education, Culture, Sports, Science and Technology of Japan.

References

Arumuganathan K, Earle ED (1991) Nuclear DNA content of some important plant species. Plant Mol Biol Rep 9: 208-218

Beguiristain T, Grandbastien MA, 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

Clark SE, Williams RW, Meyerowitz EM (1997) The *CLAVATA1* gene encodes a putative receptor kinase that controls shoot and floral meristem size in Arabidopsis. Cell 89: 575-585

Felsenstein J (1985) Confidence limits on phylogenies: an approach using the bootstrap. Evolution 38: 783-791

Felsenstein J (1993) PHYLIP (Phylogeny Inference Package) version 3.5c. Distributed by the author. Seattle, Wash.: University of Washington Department of Genetics.

Frohman MA, Dush MK, Martin GR (1988) Rapid production of full-length cDNA from rare transcripts: Amplification using a single gene-specific oligonucleotide primer. Proc Natl Acad Sci USA 85: 8998-9002

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

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

Hayakawa T (1997) Methods of DNA and RNA extraction from tobacco (*Nicotiana tabaccum*). In: Shimamoto I. Sasaki T (eds.) Protocols of PCR Experiments for Plants. Shujunsha Co, Ltd, Tokyo, pp. 45-47

Hirochika H (1993) Activation of tobacco retrotransposons during tissue culture. EMBO J 12: 2521-2528

Hirochika H (2001) Contribution of the *Tos17* retrotransposon to rice functional genomics. Curr Opin Plant Biol 4: 118-122

Hirochika H, Fukuchi A, Kikuchi F (1992) Retrotransposon families in rice. Mol Gen Genet 233: 209-216

Hirochika H, Sugimoto K, Otsuki Y, Tsugawa H, Kanda M (1996) Retrotransposons of rice involved in mutations induced by tissue culture. Proc Natl Acad Sci USA 93: 7783-7788

Ivashuta S, Naumkina M, Gau M, Uchiyama K, Isobe S, Mizukami Y, Shimamoto Y (2002) Genotype-dependent transcriptional activation of novel repetitive elements during cold acclimation of alfalfa (*Medicago sativa*). Plant J 31: 615-627

Jiang N, Bao Z, Zhang X, Hirochika H, Eddy SR, McCouch SR, Wessler SR (2003) An active DNA transposon family in rice. Nature 421: 163-167

Johns MA, Mottinger J, Freeling M (1985) A low copy number, *copia*-like transposon in maize. EMBO J 8: 1093-1102

Kimura M (1980) A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. J Mol Evol 16: 111-120

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

Kumekawa N, Ohtsubo H, Horiuchi T, Ohtsubo E (1999) Identification and characterization of novel retrotransposons of the gypsy type in rice. Mol Gen Genet 260: 593-602

Lall IPS, Maneesha, Upadhyaya KC (2002) *Panzee*, a *copia*-like retrotransposon from the grain legume, pigeonpea (*Cajanus cajan* L.). Mol Genet Genomics 267: 271-280

Laten HM, Majumdar A, Gaucher EA (1998) *SIRE-1*, a *copia/Ty1*-like retroelement from soybean, encodes a retroviral envelope-like protein. Proc Natl Acad Sci USA 95: 6897-6902

Li Q, Hunt AG. (1997) The polyadenylation of RNA in plants. Plant Physiol 115: 321-325

Manninen I, Schulman AH (1993) *BARE-1*, a *copia*-like retroelement in barley (*Hordeum vulgare* L.). Plant Mol Biol 22: 829-846

Melayah D, Bonnivard E, Chalhoub B, Audeon C, Grandbastien MA (2001) The mobility of the tobacco Tnt1 retrotransposon correlated with its transcriptional activation by fungal factors. Plant J 28: 159-168.

Murashige T, Skoog F (1962) A revised medium for rapid growth and bioassays with tobacco tissue cultures. Physiol Plant 15: 474-497

Preston BD (1996) Error-prone retrotransposition: rime of the ancient mutators. Proc Natl Acad Sci USA 93: 7427-7431

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

Shimonishi K, Fujita E, Kouzuma M, Ichi K, Ikeda K, Torigoe H (2001) Problems in sweetpotato production related to 'cultivar decline' in Kagoshima Prefecture. In: Proceedings of International Workshop on Sweetpotato Cultivar Decline Study. Kyushu Natl Agri Exp Station, Miyazaki, JAPAN, pp. 33-41

Suoniemi A, Narvanto A, Schulman A (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, Otsuki H, 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

Tanaka M, Yoshinaga M, Nakazawa Y, Nakatani M (2001) Identification of a novel SINE-like sequence in retrotransposon fragments of sweet potato. Plant Biotechnology 18: 157-161

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

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

Waugh R, McLean K, Flavell AJ, Pearce SR, Kumar A, Thomas BBT, Powell W (1997) Genetic distribution of Bare-1-like retrotransposable elements in the barley genome revealed by sequence-specific amplification polymorphisms (S-SAP). Mol Gen Genet 253: 687-694

Yamamoto E, Karakaya HC, Knap HT (2000) Molecular characterization of two soybean homologs of *Arabidopsis thaliana CLAVATA1* from the wild type and fasciation mutant. Biochim Biophys Acta 1491: 333-340

Tables

No tables are used in this manuscript.

Figure legends

Figure 1. Grouping of RTase sequences of Ty1-*copia* retrotransposons in the sweetpotato genome.

(a) Neighbor-joining tree of 50 DNA sequences of sweetpotato Ty1-*copia* retrotransposons based on the reverse transcriptase (RTase) domain inside of two conserved motifs, DVKTAFL and YVDDM.

DDBJ/EMBL/GenBank accession numbers are used to represent sequences. Numbers on the branches indicate the bootstrapping probability of 1,000 replicates. Four branches with five or more sequences show probabilities over 99%, indicating four presumably related groups of retrotransposons **(**Gr1, Gr2, Gr3 and Gr4**)**.

(b) The consensus DNA sequences of the RTase domain inside of the two conserved motifs, DVKTAFL and YVDDM, for each group of sweetpotato Ty1-*copia* retrotransposons. Underlined bold letters indicate primer sequences used for selective amplification of Gr1 retrotransposons in 3'RACE. IUPAC (International Union of Pure and Applied Chemistry) codes were used to represent nucleic acids. Dashes indicate gaps inserted to maintain alignment.

Figure 2. A schematic representation of the sweetpotato Ty1-*copia* retrotransposon, *Rtsp-1* and aligned sequences of two 3'RACE products and a transposed copy of the *Rtsp-1*.

(a) A schematic representation of the sweetpotato Ty1-*copia* retrotransposon, *Rtsp-1*. Pr, Int, RT, and RH stand for protease, integrase, reverse transcriptase, and RNaseH, respectively. The bold line A and B correspond to the 5' long terminal repeat (LTR) with the primer binding site (PBS) region and the fragment from which 3' RACE products were amplified, respectively.

(b) DNA sequence of the 5' LTR and PBS of the *Rtsp-1*.

LTR sequences are shadowed. The PBS of 13 bp follows the LTR with two intervening nucleotides. The sequences for the putative TATA and the PBS are underlined. The priming sites for the Ext_R and PBS_Comp primers are boxed.

(c) Aligned sequences of two 3'RACE products and a transposed copy of the *Rtsp-1*.

Plus signs and asterisks indicate sequence identity to the transposed copy sequence of 3' RACE-1 and -2 products, respectively. Nucleotides that differ from the transposed copy sequence are shown in capital letters. A dash indicates a deletion. 3' LTR sequences are shadowed. The group 1 (Gr1), nested Gr1 downstream and RNaseH primers are boxed. Dot blotting probe sequences of the 3'RACE-1 are underlined as well as the polypurin tract (PPT) of the *Rtsp-1*.

Figure 3. Sequence-specific amplification polymorphism (S-SAP) fluorographs produced with *Rtsp-1* LTR sequences from two callus lines (KB-1 and KB-4) and the sweetpotato "Koukei 14" plant, from which the two callus lines were derived.

Fragment size indications are based on sequencing reaction mixtures of a DNA fragment with known sequence, which were electrophoresed on the same gel. Arrows indicate polymorphic bands. The 624 bp band present only in KB-4 was cloned and sequenced.

Figure 4. PCR analysis of the KB-4 specific S-SAP product.

Genomic DNAs from a sweetpotato "Koukei 14" plant (lane 1) and two "Koukei 14"-derived callus lines (KB-1, lane 2; KB-4, lane 3) were used as templates with primers corresponding to the host *CLAVATA-1* homolog (CL1 primer) and *Rtsp-1* LTR (U3 primer) sequences as determined in the KB-4-specific 624 bp S-SAP product (compare Fig. 5b). An expected 328 bp fragment was observed only in KB-4 (lane 3), suggesting that this S-SAP site was generated by *Rtsp-1* transposition during tissue culture of KB-4.

Figure 5. PCR analysis of the *Rtsp-1* transposition in the KB-4 callus.

Genomic DNAs from a sweetpotato "Koukei 14" plant (lane 1) and two "Koukei 14"-derived callus lines (KB-1, lane 2; KB-4, lane 3) were used as templates with the *CLAVATA-1* homolog primers, CL1 (filled arrowheads) and CL1_Anti (open arrowheads), which correspond to the sequences flanking the *Rtsp-1* copy transposed during tissue culture of KB-4 callus.

(a) PCR results and a schematic representation of *Rtsp-1* insertion into the *CLAVATA-1* homolog of sweetpotato genome. The *Rtsp-1* was transposed into the second *CLAVATA-1* homolog of the KB-4 callus at the site indicated by vertical arrow. PCR based on the flanking sequences to the *Rtsp-1* insertion amplified products of 241 bp and 4,791 bp in "Koukei 14" and KB-1. The former is the *CLAVATA-1* homolog fragment without *Rtsp-1* insertion, while the latter is a fragment containing parts of two tandem aligned *CLAVATA-1*-like genes. The same PCR amplified a 5,214 bp product in KB-4, which corresponds to *Rtsp-1* (4,968 bp) inserted into the 241 bp *CLAVATA-1* homolog fragment, with 5 bp added by target site duplication. The band with the size over 10kbp appeared on KB-1 and KB-4. This indicates the possibility of three *CLAVATA-1* genes tandem aligned. The absence of the band for the Koukei 14 plant probably is due to the quality of extracted DNA for PCR amplification. High weight DNA can be easily extracted from soft callus tissue compared to leaf tissue.

(b) Aligned sequences of the *CLAVATA-1* homolog of 4,791 bp (*CLAVATA* homolog), the KB-4 specific S-SAP products of 624 bp (S-SAP 624 bp), and the *CLAVATA-1* homolog of 5,214 bp which contains a transposed *Rtsp-1* copy (*Rtsp-1* in *CLAVATA*). For clarity, 4,400 bp at the 5' end and an inner portion of 4,654 bp are omitted from the *CLAVATA-1* homolog and *Rtsp*-1, respectively. The *Mse*I recognition site, TTAA, is shown in a shadowed box. This site is followed by A, which matches with the MseI+A

selective primer sequence for S-SAP amplification. LTR sequences are shadowed and the duplicated target site sequence is boxed. Underlined bold letters correspond to the PCR primers. Dashes indicate gaps inserted to maintain alignment.

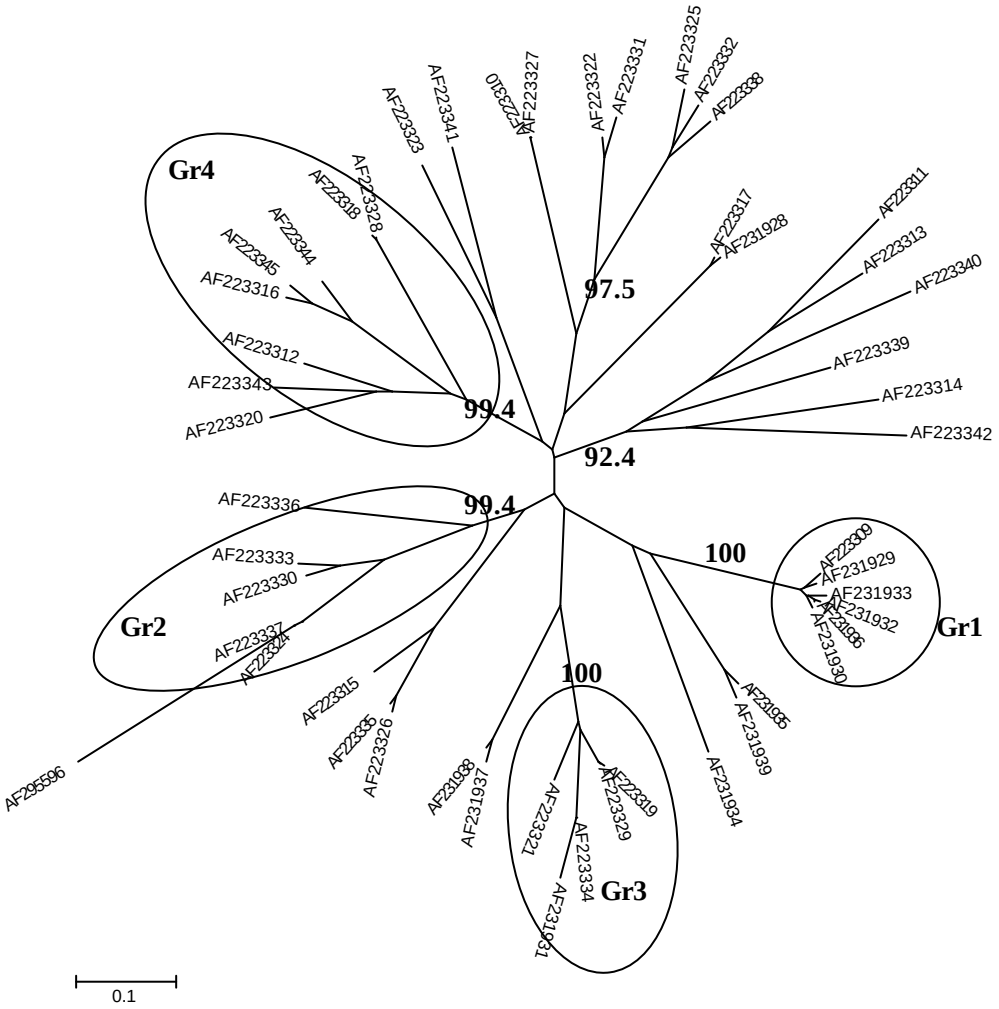

Figure 1. Grouping of RTase sequences of Ty1-*copia* retrotransposons in the sweetpotato genome.

(a) Neighbor-joining tree of 50 DNA sequences of sweetpotato Ty1-*copia* retrotransposons based on the reverse transcriptase (RTase) domain inside of two conserved motifs, DVKTAFL and YVDDM. DDBJ/EMBL/GenBank accession numbers are used to represent sequences. Numbers on the branches indicate the bootstrapping probability of 1,000 replicates. Four branches with five or more sequences show probabilities over 99%, indicating four presumably related groups of retrotransposons **(**Gr1, Gr2, Gr3 and Gr4**)**.

```
                                                    Gr1 specific primer
Gr1  TCATGGAGATCTTGAAGAAGAGATATATATGCTCCAGCCTGAAGGATTTGAAKACAAAGA
Gr2  TAAYGGTGATCTAGAAGAGGWGGTTTWTATGAAACAACCTGAAGGWTTCTCTTCTAGKRA
Gr3  RAATGRYGARTTAGAKGARGAARTSTAYRTRRAVCAACSHGARGRBYTYRTKRTTBYDGG
Gr4  HCAYGGBGAYYTKNHNGARGARGTBT----GCAACCWCCTG---GGTTYGTTGCTCAGGG

                                        Gr1 downstream primer
Gr1  GAATCAGAACTYGGTTTGYAGGTTGAACAAATCTCTRTACGGTYTAAAGCAGGCGCCAAG
Gr2  WGRYRAKMAYTTRGTTTGYAAGCTYAARAARTCYATWTWYGGAYTAAAACRAGCYTCCCG
Gr3  SCAAGAAMRDAAAGYDTGYAARTTRRTTAWRTYRYTHTATRKYTTRAARCAARCVCMWAA
Gr4  G---GARWMYRGBDWRTGYMRRCTDMRNMRVTCNYTRTATGGNYTRAARCARWS-YCN--

Gr1  GTGTTGGTATAAGAGATTTGATTCCTTCATCATGTGCCTTGGATRCAAYAGACTSAATGC
Gr2  VCAATGGTATATMAARTTYGAYGGRRTHATTTCATCATWTGGTTTTGYTGAAARTMYHMT
Gr3  KYARTGGCAYRRDMRDTTYGAYYAHKYRRTGYTARMYAATGKRTWKWRRATAAAYVAWWS
Gr4  ----TGGWTTGKRARRTTYAGYNVWGYRRTBDHNTGGNATGVNNMRNAGTRHDKNKGATC

Gr1  AGAYCCTTGTGCATATTTCAARAGGTYTGKARAAGATAACTTTGTTATATTGMTRYTV
Gr2  WGATMATTGTAYATAC--YARAAGRYAAGTGGGAGTAAAATWTGYTTTCTTG-TWYTM
Gr3  TGRYAARTGYRTMTACRBYAARR---WAACVRWRRAYRVBKWYGTHATWMTYTGYTTA
Gr4  AYWCDDTNYTHDBYRBWCBARYN---VDRVHDGYRTWHWYYTDDTWRTN---------
```

Figure 1. Grouping of RTase sequences of Ty1-*copia* retrotransposons in the sweetpotato genome.

(b) The consensus DNA sequences of the RTase domain inside of the two conserved motifs, DVKTAFL and YVDDM, for each group of sweetpotato Ty1-*copia* retrotransposons. Underlined bold letters indicate primer sequences used for selective amplification of Gr1 retrotransposons in 3'RACE. IUPAC (International Union of Pure and Applied Chemistry) codes were used to represent nucleic acids. Dashes indicate gaps inserted to maintain alignment.

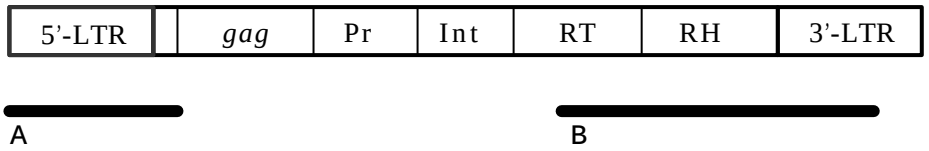

Figure 2. A schematic representation of the sweetpotato Ty1-*copia* retrotransposon, *Rtsp-1* and aligned sequences of two 3'RACE products and a transposed copy of the *Rtsp-1*.
(a) A schematic representation of the sweetpotato Ty1-*copia* retrotransposon, *Rtsp-1*. Pr, Int, RT, and RH stand for protease, integrase, reverse transcriptase, and RNaseH, respectively. The bold line A and B correspond to the 5' long terminal repeat (LTR) with the primer binding site (PBS) region and the fragment from which 3' RACE products were amplified, respectively.

TGTCGGCAAATGGAAGTTTGGCAGCAATCCATGCAAGGCAAATACGGCAGCAGCAACTTT
TGAATTAGGATGTTAGGCAACCAAATTGTGTTGACAATTATAATTATCATTTGGTCCCTA
CATTTTAGACTAACTACAAAATGACCCCAAATCTACAACTATAAATAGGGTGGTCATTTG
TCATTCTAGTCATCCCAAATCATTCTATTCTTCCCTCATATACTAGAGATATTAGAGAGT
TTGTTGAGTGTATTTCTCCTTCCTAGCAATAGAGAATTATCCTTGTTTTCTCCTTGTTAG
TTAGAGAGTGGTTGTAACTCCTATTTTCTCATAGTGAAATTCTTCTACCCTTGCCCGTGG
TTTTTACCCTAATTTGTTTAGGGGTTTTCCACGTAAAATCTGTGCCTCATTTATTTTTAC
TTTCTCAAATTATTGTTGCAATCAGATCTATCCCACTTCCGCGGGCGCAACAATTGGTAT
CAGAGCCTTCTG

Figure 2. A schematic representation of the sweetpotato Ty1-*copia* retrotransposon, *Rtsp-1* and aligned sequences of two 3'RACE products and a transposed copy of the *Rtsp-1*.

(b) DNA sequence of the 5' LTR and PBS of the *Rtsp-1*.

LTR sequences are shadowed. The PBS of 13 bp follows the LTR with two intervening nucleotides. The sequences for the putative TATA and the PBS are underlined. The priming sites for the Ext_R and PBS_Comp primers are boxed.

Figure 2. A schematic representation of the sweetpotato Ty1-*copia* retrotransposon, *Rtsp-1* and aligned sequences of two 3'RACE products and a transposed copy of the *Rtsp-1*.

(c) Aligned sequences of two 3'RACE products and a transposed copy of the *Rtsp-1*.

Plus signs and asterisks indicate sequence identity to the transposed copy sequence of 3' RACE-1 and -2 products, respectively. Nucleotides that differ from the transposed copy sequence are shown in capital letters. A dash indicates a deletion. 3' LTR sequences are shadowed. The group 1 (Gr1), nested Gr1 downstream and RNaseH primers are boxed. Dot blotting probe sequences of the 3'RACE-1 are underlined as well as the polypurin tract (PPT) of the *Rtsp-1*.

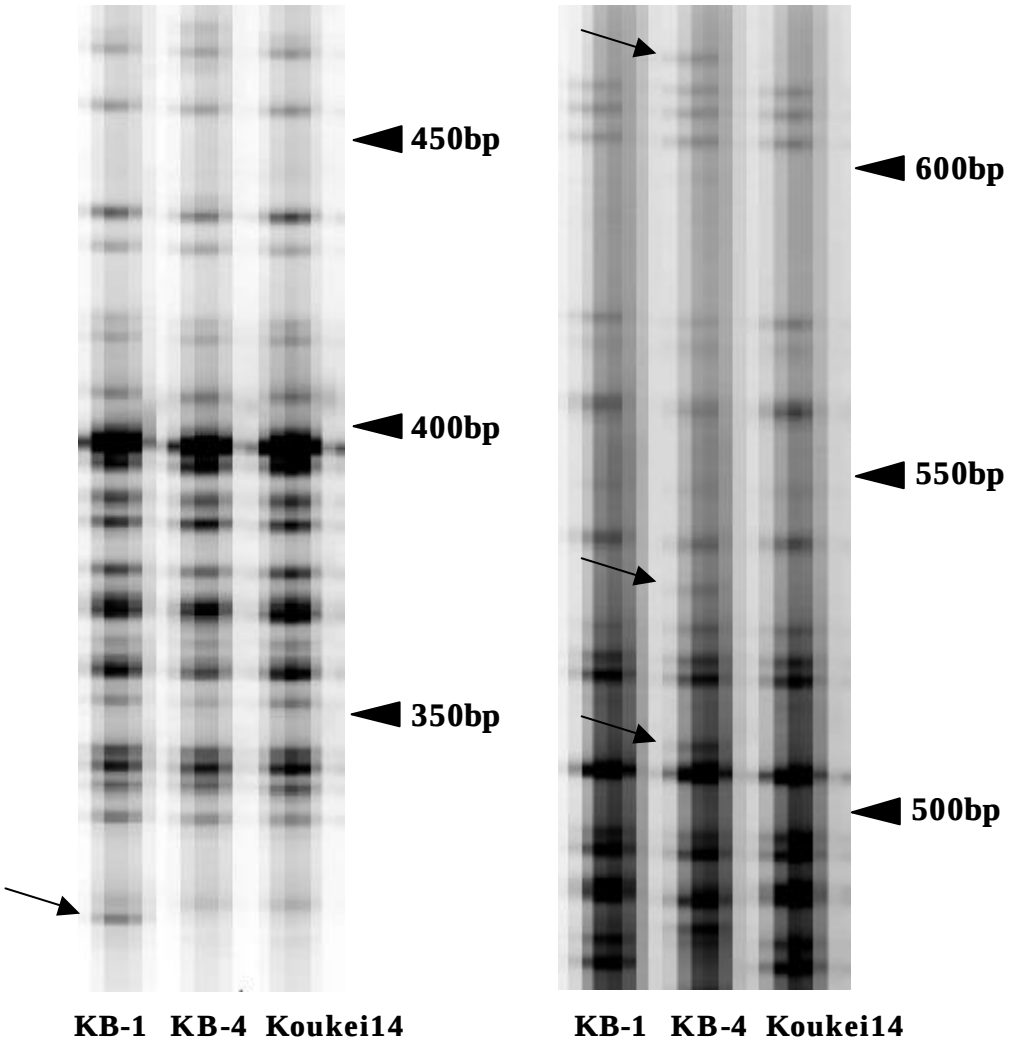

Figure 3. Sequence-specific amplification polymorphism (S-SAP) fluorographs produced with *Rtsp-1* LTR sequences from two callus lines (KB-1 and KB-4) and the sweetpotato "Koukei 14" plant, from which the two callus lines were derived.

Fragment size indications are based on sequencing reaction mixtures of a DNA fragment with known sequence, which were electrophoresed on the same gel. Arrows indicate polymorphic bands. The 624 bp band present only in KB-4 was cloned and sequenced.

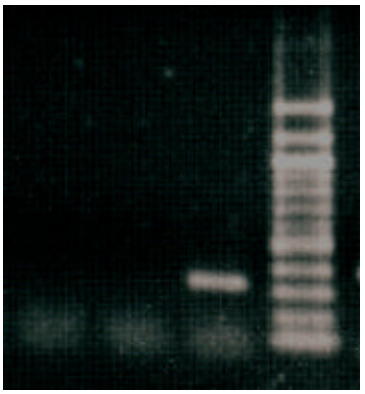

Figure 4. PCR analysis of the KB-4 specific S-SAP product.
Genomic DNAs from a sweetpotato "Koukei 14" plant (lane 1) and two "Koukei 14"-derived callus lines (KB-1, lane 2; KB-4, lane 3) were used as templates with primers corresponding to the host *CLAVATA-1* homolog (CL1 primer) and *Rtsp-1* LTR (U3 primer) sequences as determined in the KB-4-specific 624 bp S-SAP product (compare Fig. 5b). An expected 328 bp fragment was observed only in KB-4 (lane 3), suggesting that this S-SAP site was generated by *Rtsp-1* transposition during tissue culture of KB-4.

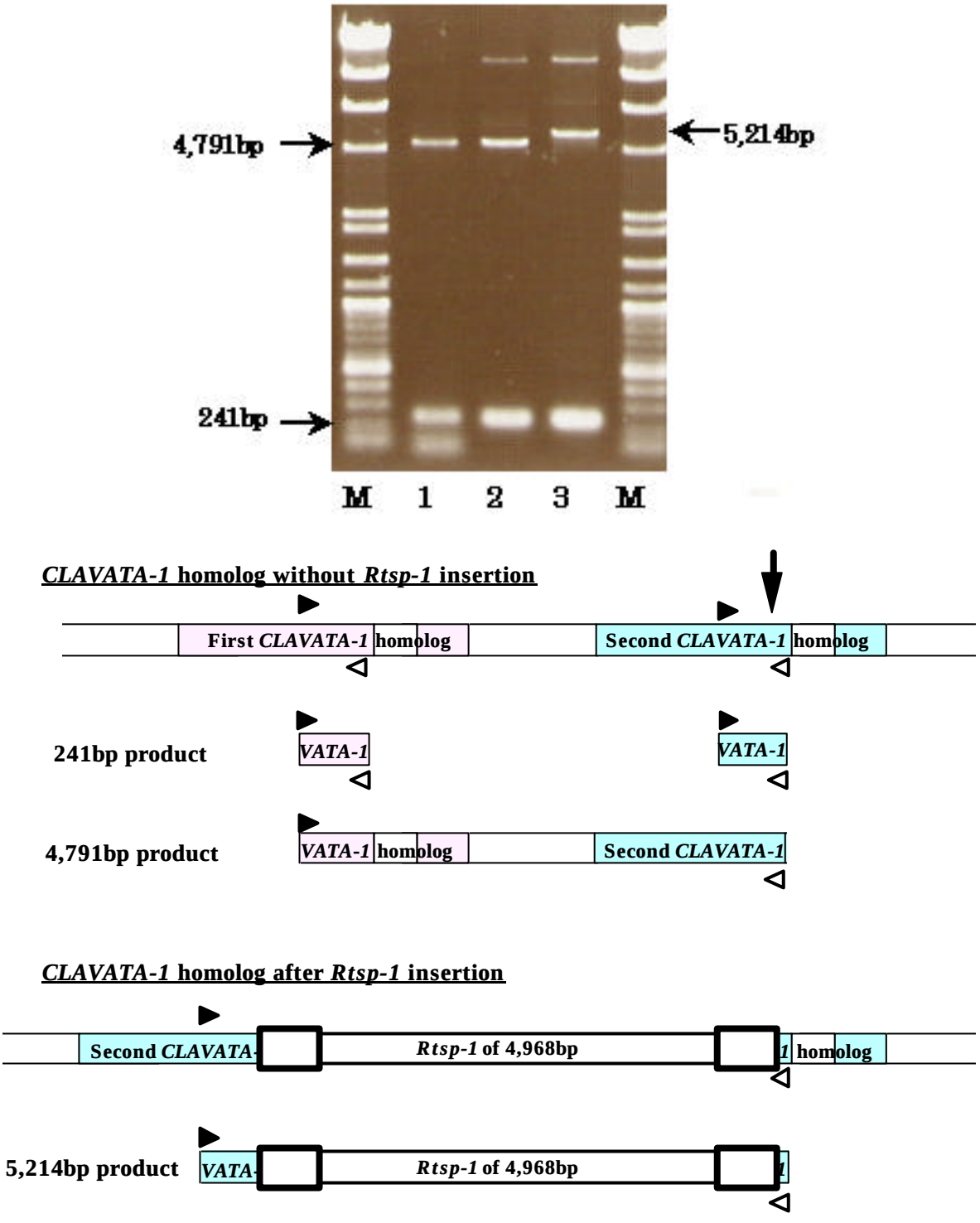

Figure 5. PCR analysis of the *Rtsp-1* transposition in the KB-4 callus.
Genomic DNAs from a sweetpotato "Koukei 14" plant (lane 1) and two "Koukei 14"-derived callus lines (KB-1, lane 2; KB-4, lane 3) were used as templates with the *CLAVATA-1* homolog primers, CL1 (filled arrowheads) and CL1_Anti (open arrowheads), which correspond to the sequences flanking the *Rtsp-1* copy transposed during tissue culture of KB-4 callus.

(a) PCR results and a schematic representation of *Rtsp-1* insertion into the *CLAVATA-1* homolog of sweetpotato genome. The *Rtsp-1* was transposed into the second *CLAVATA-1* homolog of the KB-4 callus at the site indicated by vertical arrow. PCR based on the flanking sequences to the *Rtsp-1* insertion amplified products of 241 bp and 4,791 bp in "Koukei 14" and KB-1. The former is the *CLAVATA-1* homolog fragment without *Rtsp-1* insertion, while the latter is a fragment containing parts of two tandem aligned *CLAVATA-1*-like genes. The same PCR amplified a 5,214 bp product in KB-4, which corresponds to *Rtsp-1* (4,968 bp) inserted into the 241 bp *CLAVATA-1* homolog fragment, with 5 bp added by target site duplication. The band with the size over 10kbp appeared on KB-1 and KB-4. This indicates the possibility of three *CLAVATA-1* genes tandem aligned. The absence of the band for the Koukei 14 plant probably is due to the quality of extracted DNA for PCR amplification. High weight DNA can be easily extracted from soft callus tissue compared to leaf tissue.

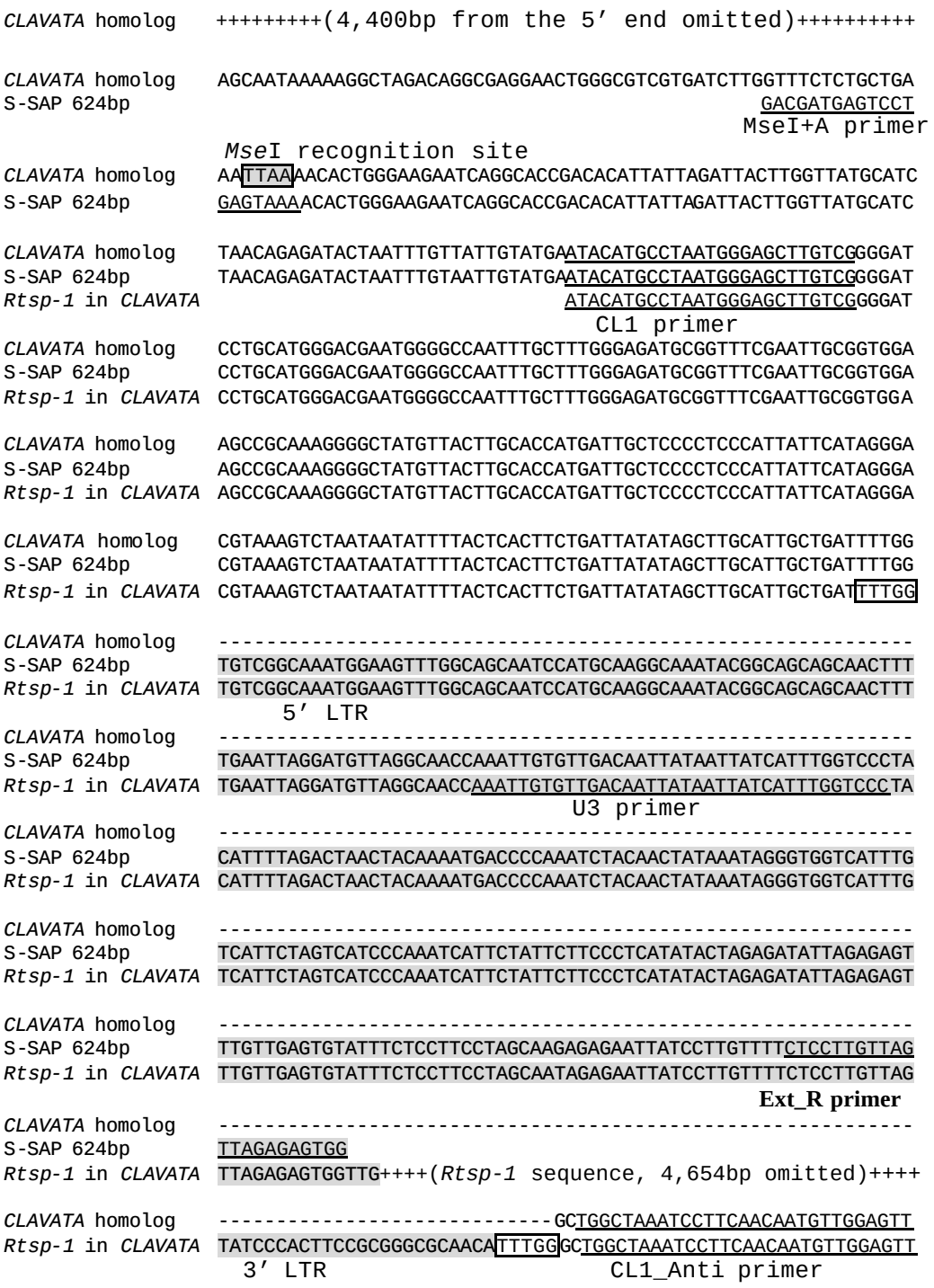

Figure 5. PCR analysis of the *Rtsp-1* transposition in the KB-4 callus.

Genomic DNAs from a sweetpotato "Koukei 14" plant (lane 1) and two "Koukei 14"-derived callus lines (KB-1, lane 2; KB-4, lane 3) were used as templates with the *CLAVATA-1* homolog primers, CL1 (filled arrowheads) and CL1_Anti (open arrowheads), which correspond to the sequences flanking the *Rtsp-1* copy transposed during tissue culture of KB-4 callus.

(b) Aligned sequences of the *CLAVATA-1* homolog of 4,791 bp (*CLAVATA* homolog), the KB-4 specific S-SAP products of 624 bp (S-SAP 624 bp), and the *CLAVATA-1* homolog of 5,214 bp which contains a transposed *Rtsp-1* copy (*Rtsp-1* in *CLAVATA*). For clarity, 4,400 bp at the 5' end and an inner portion of 4,654 bp are omitted from the *CLAVATA-1* homolog and *Rtsp*-1, respectively. The *Mse*I recognition site, TTAA, is shown in a shadowed box. This site is followed by A, which matches with the MseI+A selective primer sequence for S-SAP amplification. LTR sequences are shadowed and the duplicated target site sequence is boxed. Underlined bold letters correspond to the PCR primers. Dashes indicate gaps inserted to maintain alignment.