Available online at www.sciencedirect.com

SCIENCE DIRECT®

Virology 332 (2005) 614–622

VIROLOGY

www.elsevier.com/locate/yviro

Occurrence of a DNA sequence of a non-retro RNA virus in a host plant genome and its expression: evidence for recombination between viral and host RNAs

Edna Tanne[1], Ilan Sela*

Virus Laboratory, The Hebrew University of Jerusalem, Faculty of Agricultural, Food and Environmental Quality Sciences, Robert H. Smith Institute for Plant Sciences and Genetics, Rehovot 76100, Israel

Received 8 September 2004; returned to author for revision 26 October 2004; accepted 9 November 2004
Available online 8 January 2005

Abstract

This study demonstrates that sequences homologous to those of the non-retro RNA virus (Potato virus Y; PVY) are integrated into the genome of several grapevine varieties. The integrated PVY-coat-protein-like cistron is expressed in the grapevine as indicated by Southern and Western blot analyses as well as by RNase protection assay. In addition, genome-walking studies showed that one PVY-like sequence is flanked by 41-bp direct repeats and is embedded in authentic grapevine sequences, flanked by inverted repeats. Rearranged PVY-like sequences were also found in tobacco. It is suggested that nonhomologous recombination of a potyviral RNA with RNA of a retrotransposable element took place at some point in evolution. The initial integration locus was probably within a grapevine gene homologous to a pentatricopeptide repeat-carrying protein, and was later transposed to other locations. The current location is reminiscent of a MITE-type retroelement, indicating transposition history. Because grapevine cultivars are propagated asexually, without going through a meiotic phase, the chance for DNA recombination is minimal and the foreign integrated sequence may be better conserved, enabling it to be expressed correctly in the recipient genome.
© 2004 Elsevier Inc. All rights reserved.

Keywords: PVY; Integration; Grapevine; Transposition; RNA recombination

Introduction

Viral DNA sequences can become incorporated in host genomes, and partial sequences of viral DNAs have been found integrated in those genomes. DNAs of pararetroviruses have been found in the host nucleus, either integrated into the host DNA, or in the form of episomes (Jakowitsch et al., 1999; Lockhart et al., 2000; Mette et al., 2002; Ndowora et al., 1999; Richert-Poggeler et al., 2003). Sequences of the single-stranded DNA geminiviruses have also been found integrated in their host plant's genome (Ashby et al., 1997; Bejarano et al., 1996; Harper et al., 1999). However, the only viruses with RNA genomes known to integrate a DNA version of their genome into host chromosomal DNA are the retroviruses. In this case, the RNA genomes are reverse-transcribed and the resultant DNA is inserted into the host DNA by a virus-encoded integrase; these reactions are required for normal replication (for example, Flint et al., 2004; Goff, 1992; Hu and Temin, 1990).

Potyviruses are a large, polyphagous group of plant viruses that carry an RNA genome of sense orientation. Potyviral RNA is replicated via an RNA replicase, and does not undergo reverse transcription or genome integration (Lazarowitz, 2001). Extracts of several (but not all) grapevine (*Vitis vinifera*) varieties have been reported to react with antiserum against the potyvirus potato virus Y (PVY). Dot

* Corresponding author. Fax: +972 8 9473402.
E-mail address: sela@agri.huji.ac.il (I. Sela).
[1] Permanent address: Department of Virology, Agricultural Research Organization, Volcani Center, Bet Dagan 50250, Israel.

0042-6822/$ - see front matter © 2004 Elsevier Inc. All rights reserved.
doi:10.1016/j.virol.2004.11.007

blot analysis of nucleic acids extracts also reacted positively with a potyviral probe (Naveh, 1987; Tanne et al., 1989). However, potyvirus particles or potyviral-associated inclusion bodies (such as pinwheels) have never been found in the grapevines. Here we demonstrate that grapevine genomic DNA carries the entire gene of a potyviral coat protein (CP) and the potyviral 3′ UTR. Potyviral-homologous sequences were also found in tobacco DNA, albeit in a rearranged form.

Transposable elements (TE) are divided into two general classes. Class I elements are propagated by a mechanism that involves RNA intermediates. Some active Class I elements (retroviruses and retrotransposons) carry long terminal repeats (LTRs) at both ends. Such elements typically encode their own reverse transcriptase and integrase, allowing the production of a DNA copy from an RNA transcript and its insertion at another site in the genome. Other types of class I elements lack LTRs and are either transposed by another mechanism (LINES) or do not encode transposition proteins (SINES) but depend on those of other elements, such as LINES. Class I elements are flanked on either side by direct repeats of host sequences of various length, formed at the time of their integration into the host DNA.

Class II TEs are mobilized by a "cut and paste" mechanism. Active class II transposons carry terminal inverted repeats (TIRs), and are also flanked by direct repeats of chromosomal sequences. These transposons encode one or two open reading frames (ORFs), one of which specifies a transposase, responsible for mobilization. Many types of inactive (nonautonomous) class II elements have evolved, comprising an array of repetitive sequences dispersed throughout the host genome. One such group, miniature inverted-repeat transposable elements (MITEs), has been described recently in plants and animals, where it is present in a very high copy number (for example: Fechotte and Mounchės, 2000; Le et al., 2000; Santiago et al., 2002; Zhang et al., 2001). MITEs are TIR-carrying short dispersed sequences (<500 bp). These elements do not carry any ORF; however, they are located next to transcribed sequences (Wessler et al., 1995; Yang et al., 2001). Sequence analysis has indicated that MITEs have probably evolved from active class II transposons, as some degree of homology was found with transposase-carrying elements (Fechotte and Mounchės, 2000; Le et al., 2000). In maize, a class II group of size range 358 bp to 5.2 kbp, called *P instability factors* (*PIF*), comprise several nonautonomous members and one active element (*PIFa*; Zhang et al., 2001). MITEs are structurally similar to the defective PIF elements. However, their high copy number suggests a replicative mechanism of propagation and transposition (Fechotte and Mounchės, 2000).

The data presented in this paper suggest that recombination between viral RNA and the RNA of a host cell retrotransposable element may have taken place. RNA recombination between various segments of viral RNAs is a well-recorded phenomenon, and various mechanisms have been proposed to account for these recombination events (Nagy and Simon, 1997, and references therein). However, all known recombination events occurred between segments of RNAs of the same virus, or related viruses (some recent examples: Adams et al., 2003; Oberste et al., 2004; Wu et al., 2003). A major mechanism leading to viral RNA recombination appears to be template switching by the viral replicase (Nagy and Simon, 1997). Indeed, mutation in the viral replicase is reported to reduce RNA recombination (Cheng and Nagy, 2003). This may explain the restriction of RNA recombination to segments of RNAs of the same virus or between "strains" of the same virus, as these RNAs are all recognized by the same replicase. A similar mechanism is also suggested for the appearance of defective interfering viral RNAs (e.g., Bar-Joseph et al., 1997; Shapka and Nagy, 2004). AU rich regions, and RNA promoters for subgenomic RNAs seem to have a role in RNA recombination (Shapka and Nagy, 2004), and host factors also influence recombination (Dzianott and Bujarski, 2004). Splicing and processing of RNAs (such as of ribosomal RNAs) are also forms of RNA recombination, in which case parts of the very same transcripts are ligated together. Indications of non-replicative viral RNA recombination (Gallei et al., 2004) and virus-host RNA recombination (Charini et al., 1994; Meyers et al., 1989; Monroe and Schlesinger, 1983) have been reported. To the best of our knowledge, recombination between host derived retrotransposable elements and (non-retro) viral RNAs has not yet been reported.

Results

The data presented in Table 1 support previous findings (Naveh, 1987; Tanne et al., 1989) by showing that several grapevine varieties reacted in an ELISA test with an antiserum against PVY. Previously, we demonstrated that nucleic acid extracts from several grapevine varieties react with a PVY probe upon dot-blot hybridization assays (Tanne et al., 1989). The absence of any potyvirus particles in the plants analyzed suggested the possible occurrence of PVY-homologous sequences in the host genome. We

Table 1
ELISA results following reaction of saps from several grapevines with antiserum to PVY

Grapevine variety or clone	ELISA reading A_{405}
LN33	0.032
Carignian	1.062
LN_J4	1.160
Rouge de Loire	0.535
Superior	1.334
Pinotage	0.510
Mission	0.017
Sultanina	0.102
Gammay	0.507
Chasselas Pully	0.080

therefore first screened DNA extracts from various varieties by PCR, amplifying the 3′-end section of the potyviral genome containing the CP gene and the 3′ UTR with primers 1 and 2 (Table 2). The expected fragment (1165 bp) was obtained from DNA templates of several grapevine varieties. Similar PCR assays with tobacco (*Nicotiana tabacum* var. Samsun NN) DNA using primers 3 and 4 (Table 2) resulted in the amplification of a shorter (278 bp) PCR products (Fig. 1).

The most consistently positive reaction was obtained with DNA extracted from the grapevine variety "Superior." The amplified "Superior" product contained a sequence strikingly similar to that of the 3′ section of PVY, including sequences homologous to PVY-CP and PVY-3′ UTR (compared to the PVY sequence in GenBank accession no. D00441) and exhibiting an uninterrupted reading frame with predicted amino acid sequence similarity to PVY-CP protein (Fig. 2A). The sequence of one short tobacco band included a 99-bp segment homologous to PVY; however, in this case, the potyviral sequence is rearranged and the CP gene sequence is interrupted (Fig. 2B). This is too short a sequence; therefore, we refrain from drawing conclusions at

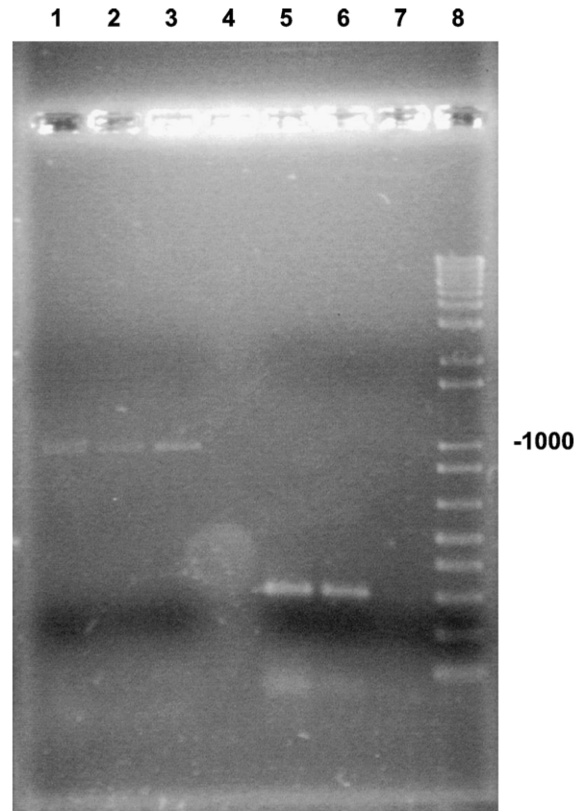

Fig. 1. Electropherogram of PCR products amplified from grapevine and tobacco DNAs with primers designed for PVY. The DNA-templates for PCR in the various lanes were extracted from the following grapevine varieties: Gammay (lane 1), Pinotage (lane 2), Superior (lane 3), LN33 (lane 4), and from tobacco Samsun NN (lanes 5 and 6). Lane 7 represents a negative control (without a template). Size markers are shown in lane 8. The number on the right indicates the position of a size marker of 1000 bp.

this point, especially since we could not yet find a tobacco clone with the PVY sequences.

To corroborate the PCR results and ensure that the potyviral sequence is genome-integrated, we performed Southern blot analysis with a segment of a cloned PVY-CP as a probe. To check for possible expression of the inserted potyviral gene segment, we employed RNase protection assays (with this PVY-derived probe) and Western blot assays with antibodies to PVY (Figs. 3–5). These assays indicated the occurrence in grapevine of DNA, RNA, and protein sequences similar to those of a potyvirus. Analysis of the obtained integrated sequence (described further on) indicated that the resultant host–viral fused protein carries 29 amino acids of host origin at its N-terminus. On the other hand, 22 N-terminal amino acids of viral origin were deleted. Consequently, the product of the integrated sequence differed from PVY-CP by seven amino acids, which is beyond the resolution of a gel detecting proteins of ca. 30 kDa, and the fused protein therefore migrated to a position similar to that of PVY-CP.

To further corroborate the integration of the potyviral sequences and determine the site of integration in the *Vitis*

Table 2
List of primers used in PCR assays and genome walking

No.	Primer designation*	Sequence of primer
1	Forward primer for amplifying the coat_protein and 3′ UTR section of PVY	5′ CAG CCA AAC CCG AAC AAA G 3′
2	Reverse primer for amplifying the coat_protein_3′ UTR section of PVY	5′ CTA ACC CGG GTC TCC TGA TTG AAG TTT ACA GTC 3′
3	Forward primer for amplifying the coat_protein_3′ UTR section of PVY	5′ ACT GTG CCG AGA ATC AAG 3′
4	Reverse primer for amplifying the coat_protein_3′ UTR section of PVY	5′ CAT CAT AAC CCA AAC TCC 3′
5	Gene walk primer to extend downstream of the PVY 3′ UTR	5′ GTC TGG ATT TAG TTA CTT GGG TGA TGC TG 3′
6	Gene walk primer (nested) to extend downstream of the PVY 3′ UTR	5′ GTC ATA GCA GTG ACT GTA AAC TTC AAT CAG G 3′
7	Gene walk primer to extend upstream to the PVY coat_protein gene	5′ CGG TTG CTC CCT TGC TTG TGG GCA TTC TCA 3′
8	Gene walk primer (nested) to extend upstream to the PVY coat_protein gene	5′ GGA CGT GAT AGC CTT GAT TCT CGG CAC AGT ATG 3′
9	Gene walk primer to go backward across the PVY sequence of clone GWA5	5′ GGC CCT CAC CAC CTA ATT ATG GAT GCC CTT CTC 3′
10	Gene walk (nested) primer to go backward across the PVY sequence of clone GWA5	5′ AGA AGC TAC AAA GAG CTG ATG TCA TTC CAC 3′

* Primers 1_8 were designated according to the PVY sequence in GenBank accession number D00441. Primers 9, 10 were designated according to the grapevine sequence adjacent to the potyviral sequence of clone GWA5.

A.

```
                           10        20        30        40        50
pvy-in-grape     SVPRMLQTRYVSDPEFVIQPNPNKGKDKDVNAGTSGTHTVPRIKAITSKMRMPKSK
                                  |||||||||||||||||||||||||||||||||||| ||
pvy-cp.pep   ANDTIDAGGSNKKDAKPEQGSIQPNPNKGKDKDVNAGTSGTHTVPRIKAITSKMRMPTSK
                     10        20        30        40        50        60

              60        70        80        90       100       110
pvy-in-grape GVAALNLEHLLEYTPQQIDISNTRATQSQFDTWYEAVRMAYDIGETEMPTVMNGLMVWCI
             |::: ||||||||:||||||||||||||||||||||||||||||||||||||||||||||
pvy-cp.pep   GATVPNLEHLLEYAPQQIDISNTRATQSQFDTWYEAVRMAYDIGETEMPTVMNGLMVWCI
                     70        80        90       100       110       120

              120       130       140       150       160       170
pvy-in-grape ENGTSPNVNGVWVMMDGSEQVEYPLKPIVENAKPTLRQIMAHFSDVAEAYIEMRNKKEPY
             |||||||||||||||||:||||||||||||||||||||||||||||||||||||||||||
pvy-cp.pep   ENGTSPNVNGVWVMMDGNEQVEYPLKPIVENAKPTLRQIMAHFSDVAEAYIEMRNKKEPY
                    130       140       150       160       170       180

              180       190       200       210       220       230
pvy-in-grape MPRYGLIRNLRDISLARYAFDFYEVTSRTPVRAREAHIQMKAAALKSAQPRLFGLDGGIS
             ||||||||||||::||||||||||||||||||||||||||||||||||||||||||||||
pvy-cp.pep   MPRYGLIRNLRDMGLARYAFDFYEVTSRTPVRAREAHIQMKAAALKSAQPRLFGLDGGIS
                    190       200       210       220       230       240

              240       250       260
pvy-in-grape TQEENTERHTTEDVSPSMHTLLGVKN
```

B.

TATTCCCATCCATCATAACCCTAACCCGGGTCTCCTGATTGAAGTTTACAGTCTTCTCTT
TGTTCGGGTTTGGCTGCACAGTATGTGTCCCAGAAAGCC

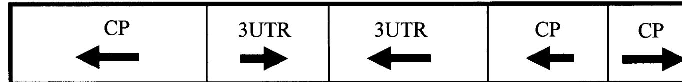

Fig. 2. (A) FASTA analysis comparing the translated sequence of the PCR product of grapevine variety Superior (pvy-in-grape) and that of the translated CP gene of the virus PVY (pvy-cp.pep). (B) Sequence of a PCR product isolated from tobacco and a schematic illustration of its homology to PVY and of rearrangement of the PVY gene sequences. Illustration is not to scale.

genome, we performed genome-walking experiments on DNA extracted from grapevine “Superior”. Primers of the potyviral CP-3′ UTR terminal sequences were extended upstream and downstream as described in Materials and methods. Several clones were found to carry the PVY sequence. In clone GWA5, the sequence upstream of PVY showed 63.5% nucleotide identity and amino acid homology (59% identity, 73% similarity along 154 amino acids) to a *Vitis* SCAR marker for a seedlessness-associated gene (grapevine variety Flame Seedless; accession no. AY327514). Primers that extended from the SCAR marker through the PVY sequence into the other end of the PVY genomic insert were used for an opposite genome-walking experiment. The obtained clone (GWB5) revealed the continuation of the grapevine SCAR sequence at the other end of the PVY (Fig. 6; data deposited with GenBank, accession no. AY615284). We conclude, therefore, that the PVY insert interrupts the *Vitis* SCAR sequence. Upon annotation of the *Vitis* SCAR marker sequence, we found TIRs at both ends of the published sequence (beyond GWB5; Fig. 7). The SCAR sequence does not extend beyond the TIRs; therefore, we could not establish if target site duplications are present. A schematic illustration is shown in Fig. 7A. Interestingly, the PVY sequence (within the SCAR locus) is flanked on both sides by identical short (41 bp) direct repeats. An open reading frame starts at the grapevine SCAR sequence and continues uninterruptedly from the grapevine sequence through the end of the PVY-CP cistron (positions 349–1300 of Fig. 6).

The 41-bp repeats were homologous (67.6%) to those present in an *Arabidopsis thaliana* gene for a protein carrying pentatricopeptide repeats (PPRs; accession no. NM_112693). In another grapevine-derived clone (GWA11; GenBank accession no. AY615285), also generated by genome walking from the PVY sequence, the PVY insert was flanked on one side (opposite walking was not performed) by a sequence homologous (71.6% identity along 820 bases) to the *Arabidopsis* PPR gene (schematically illustrated in Fig. 7B).

Blast analysis revealed that the sequence in the first 18 base pairs of the aforementioned 41-bp repeats is widely dispersed in rice, and several copies of this sequence are found in each of the following rice chromosomes: 2, 3, 4, 5, 6, 8, 9, 10, and 11. Several mammalian genes (human, apes, etc.) carry sequences homologous to the second half of the 41-bp repeats (bases 16–41, approximately, depending on

Fig. 3. Southern blot analysis of *Vitis* DNA. ^{32}P-labeled clone of PVY-CP served as a probe. Lane 1: DNA was extracted from the *vitis* cultivar Superior. Lane 2: DNA was extracted from a seedling of the *Vitis* cultivar Bonyer that did not react with antibodies to PVY in an ELISA test.

the organism). Contrary to the situation in rice, there is no match in *Arabidopsis* to the 41-bp repeats other than in the aforementioned PPR gene.

Discussion

The present study shows that potyviral-like sequences are integrated in the *Vitis* genome. This finding may lead to new avenues of thought regarding genetic diversity, the origin of viral sequences, RNA recombination and horizontal transfer of genetic material. Therefore, it was essential to determine unequivocally that the potyviral sequences were embedded in the plant genome. To that end, all DNA samples were RNase-treated, and in addition to the PCR and Southern blot analyses, we determined (by gene walking) two of the *Vitis* loci where the viral sequences had been inserted.

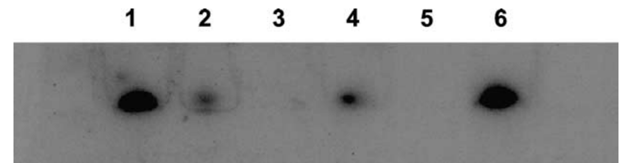

Fig. 4. RNase protection assay. RNA extracted from grapevine (cultivar Superior) was hybridized with a labeled antisense RNA fragment of PVY-CP, submitted to RNase digestion and electrophoresed. All resulting labeled bands were of the expected size (400 bases). Lanes 1, 2, and 4: RNA was extracted from three different grapevines (Superior). Lane 6: positive control. Sense-oriented RNA was transcribed from PVY-CP clone and reacted with the same antisense probe. Lanes 3 and 5 are empty.

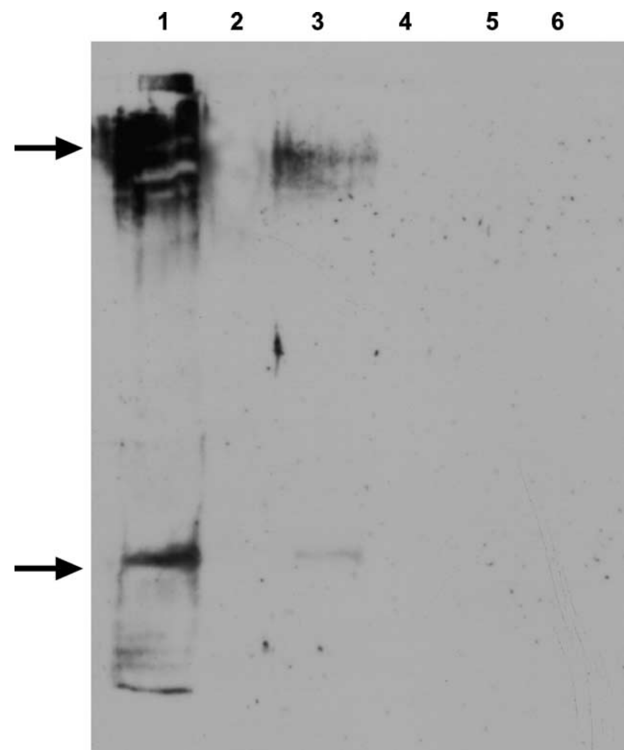

Fig. 5. Western blot analysis of grapevine proteins. Procedure was performed as described in the text. Lane 1: positive control. Proteins were extracted from PVY-infected tobacco. Lane 2: size markers (not reacting with the PVY antibodies). Lane 3: proteins extracted from the grapevine cultivar superior. Lanes 4 and 5: proteins extracted from seedlings of the grapevine cultivar Bonyer (not reacting with antibodies to PVY in an ELISA test). Lane 6: empty. The upper arrow points at a position of high molecular weight protein (beyond the resolution of the gel). The lower arrow points at the position of the soybean tripsin inhibitor (~33 kDa), approximately corresponding to the size of PVY-CP.

1 AATTCAGATT ACTATAGGGC ACGCGTGGTC GACGGCCCGG GCTGGTATCA

51 TATGGAGCAT CTGCTGGCCC CAAGAGATTT CTTCTACCCC AGAGTAGCCA

101 TGGATTTTTA TCAGTCCATG ACTACTAAAC AGGTCAGAGA TCCAACTTTA

151 ATCCATTTTA CTATAGATGG TCGGCATGGC ATTTTAGGAG CTCGCCATAT

201 AGCAGAAGCC CTGCATATAC CATACGAGCC AAGCCATTTT GAAGATTACC

251 AAGTATGGAC TAGTCCCTCT CAGCTGGAAA TGGTTCATAT CTTGTCCAGA

301 TGAGCTTCCA CAAATCCACA TCTGTTGAGG GGGGAACTTC CTCCGAGCAT

351 GTTTCTCATT GATGCACTTT TGCGTCACAA CATATATCCA CTCCAGCACT

401 GGACTCAGAG GAGAGGAGTT CTTTTGGAGG CCCTATTCAA GATTTCAGAG

451 GGATACTTCT TTGGCCCTCA CCACCTAATT ATGGATGCCC TTCTCTATTT

501 TGAAGAGAAG GTGCATAAGA **AGAAGCTACA AAGAGCTGAT GTCATTCCAC**

551 **TTCTTTTTCC A**AGACTTCAG CCAAACCCGA ACAAAGGAAA AGATAAGGAC

601 GTGAATGCTG GCACATCTGG AACACATACT GTACCGAGAA TAAAGGCTAT

651 CACGTCAAAA ATGAGAATGC CTAAAAGCAA AGGAGTGGCC GCACTAAACT

701 TGGAACACTT GCTTGAGTAT ACTCCACAAC AGATAGACAT CTCAAATACT

751 CGGGCAACTC AATCACAGTT TGATACGTGG TATGAAGCAG TGCGGATGGC

801 ATACGACATA GGGGAAACTG AAATGCCAAC TGTGATGAAT GGGCTTATGG

851 TTTGGTGCAT TGAAAATGGA ACCTCGCCAA ATGTCAACGG AGTTTGGGTT

901 ATGATGGATG GAAGTGAACA AGTTGAATAT CCGTTGAAAC CAATCGTTGA

951 GAATGCAAAA CCGACCCTTA GGCAAATCAT GGCACATTTC TCAGATGTTG

1001 CAGAAGCGTA TATAGAAATG CGCAATAAAA AGGAACCATA TATGCCACGA

1051 TATGGTTTAA TTCGAAATCT GCGGGATATA AGTTTAGCGC GCTATGCCTT

1101 TGACTTTTAT GAAGTTACAT CACGAACGCC AGTGAGGGCT AGGGAAGCGC

1151 ACATACAAAT GAAGGCCGCA GCATTAAAAT CAGCTCAACC TCGACTTTTC

1201 GGGTTGGATG GTGGCATTAG TACACAAGAG GAGAACACAG AGAGGCACAC

1251 CACCGAGGAT GTTTCTCCAA GTATGCATAC TCTACTTGGA GTCAAGAACA

1301 TGTGATTGTG ATGTCTCTCC GGACGATATA TAAGTATTTA CATATGCAGT

1351 ATGTATTATG GCTTTTCCTG TACTACTTTT ATTGCAACTA ATAATCAGTT

1401 GGATATTATT AATAAATAGA GGTGGCAGGA TGATTTCGTC ATTGGGGTGA

1451 CTCTATTTT**A GAAGCTACAA AGAGCTGATG TCATTCCACT TCTTTTTCCA**

1501 CGGCTGCTAT GTCAGATTCT GGAGCATCTG GGGTATCCAT CAGATCCTCA

1551 GTTGGAGCGC AAGCGTATAT GCCGAGAGGT ATTTACTCTC GACAAGTGGA

1601 CCAATATGAC AGCGTATAGA GTTGAGCAGC CAGGGCGCCC ACAACCAGCT

1651 GAGATACCAA CTGCCAGGAG AGCATCTCCA CATCATATAC CTGAGGGTAT

1701 ACCCATTGCT TCTCCTGCCA TACCCAGAGC TCCTGCAGTT ACTCCAGCCT

1751 CATCTGAGCC ATTTACTTCA GCTGAGTCGA GGATGGCACA CCCCCCAACC

1801 GACTCATTGC TAGTCCCTTA GCAATGGAGG AGATAAAAAC TAAGTAAACT

1851 ATAATAATAT ACATAAAAAG GATCATGCAA ATCACTCCAA AAAATAATTG

1901 AGTGGCATGA CTGATATCAA AACACATCTC ACATGATGAA CTAAAGATAC

1951 GAAGATTCAA AATGCAATAG GAGTTGTCAA AGCTAAAACT TAAACATAAA

2001 GTACAAAGAG TGGAGATTAT GCAATTAAGT ATCAAAAGAA TCTCTACTCT

2051 CAAGGTTCCG GATCAAACTC TTCCATAATG AGCTAAGTGT TAATGAGATT

2101 AGCTTCCGGA TATAGTATGC ACAACTATCC TCTCCCCTCA ACCTAAGCTT

2151 TTCTCGGACA TTGGCAAAAT TAACCAACTC TTAATAGGTT AGGAACGCAC

2201 CTTCTTATTC ACAATTCTTT TTTACTCACT AAACTTAACC AATTCACCTA

2251 TGTAGCAAGC AGAGGATCGA TGACTCCCAC CTATCAGGGT TTAGGGCATC

2301 GGGTTTTAAA GGCTTTCGCC ACCCCTTCAA

Fig. 6. The sequence of clone GWB5 showing a potyviral insert within the sequence of a grapevine SCAR marker for seedlessness and flanking direct repeats. The potyviral sequence is underlined. The 41-bp direct repeats are in bold.

Retrotransposons carry long terminal repeats (LTRs) and are flanked by host direct repeats. In many cases, such as in the Ty element of yeasts, only solo and fragmented solo LTRs are found, marking the sites of past retrotransposition events (Neuvéglise et al., 2002; Zickler and Klencker, 1999). In addition, direct repeats, of various lengths, of host sequences are created upon retrotransposition. The PVY-like insert is flanked by direct repeats, suggesting a retrotransposition history for the potyviral integration. Class II TEs carry TIRs at both ends. The PVY-like insert (bounded by direct repeats) is embedded in a SCAR marker with features characteristic of TEs. The presence of TIRs and the lack of any significant ORF may classify it as a MITE. As aforementioned, MITEs have evolved from larger TEs, and intermediate structures were found, as described for *PIF*s. The size of the SCAR-TE is larger than the conserved MITE length (100–500 bp). The expected limited number of recombination events in grapevine (discussed below) may slow the TE evolution from a classical transposon structure to MITE, and the presently described SCAR-TE is probably analogous to *PIF* intermediates. A portion of the 41-bp PPR sequence is quite abundant in rice, whereas it is scarce in *Arabidopsis.* The abundance and chromosomal location of the portions of the 41-bp repeats suggest hotspots for DNA or retro-derived recombinations.

The above data indicate that at least two copies of potyviral sequences are integrated into the "Superior" grapevine genome and, as suggested by ELISA and PCR, in some (but not all) other grapevine varieties as well. Some of the potyviral sequences have apparently remained conserved during evolution, retaining the CP open reading frame (Figs. 2 and 6), which can be expressed. Such expression can explain the fact that plants not infected with any potyvirus react to PVY antisera. Potyviruses are expressed as polyproteins undergoing processing to produce the mature viral proteins. This is exemplified in Fig. 5 (lane 1). Parts of active retrotransposons (such as *gag*) are also expressed as polyproteins undergoing further processing. In *Vitis*, a protein a bit larger than the PVY-CP (as expected from Fig. 6) reacts with the potyviral antibodies. In addition, a larger pertinent protein is detected at the top of the gel (Fig. 5, lane 3). We may presume that the ca. 33 kDa protein is translated from the sequence presented in Fig. 6 or similar sequences that may have dispersed in the *Vitis* genome. The larger protein (if not a technical artifact) may indicate the presence of a PVY-carrying transposon.

The sequence of clone GWB5 indicates that the PVY-like insert in the grapevine variety "Superior" interrupts a grapevine sequence that is homologous to a seedlessness SCAR marker, with features reminiscent of a TE, in another grapevine variety. In another clone (GWA11), the PVY sequence was found adjacent to a gene homologous to an *Arabidopsis* PPR gene, which carries sequences largely homologous to that of the 41-bp repeats. We

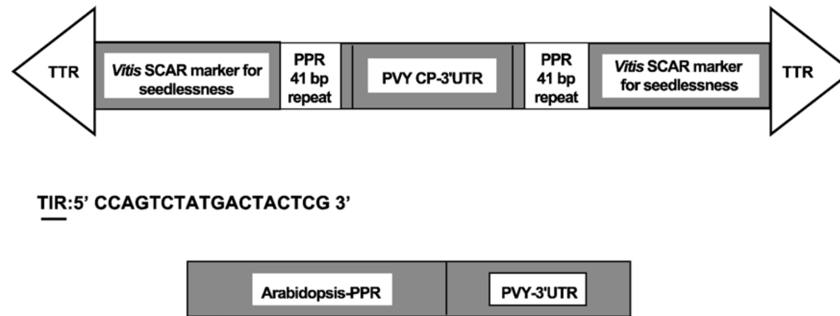

Fig. 7. (A) Schematic illustration of the sequence of clone GWB5, including *vitis* SCAR sequences beyond the clone. (B) Schematic illustration of clone GWA11. Illustrations are not to scale.

speculate that at a certain point in evolution, a potyviral sequence recombined with the RNA of a retrotransposable element, perhaps via template switching of the reverse transcriptase or the viral replicase. The chimeric DNA was subsequently integrated into the *Vitis* genome. Because one clone was found adjacent to a PPR sequence and the PVY insert in another clone is flanked by repeats of PPR homology, we speculate that the original site of integration was within a PPR-like sequence. As with many other retrotranspositions, DNA recombination events may have taken place at a later stage, and the viral insert has been further transposed to the aforedescribed SCAR location; hence the PPR homology in both clones and the TIRs at the ends of the SCAR sequence. Taken together, we tentatively suggest the following sequence of events: (i) Reverse transcription or replicase-mediated recombination took place between a retroelement and a viral RNA. (ii) The insertion of the resultant chimeric sequence into the host DNA generated the 41-bp repeats. The original (or an early) site of this insertion was within a PPR-like sequence. (iii) Further movement based on homology with the 41-bp repeats has taken place. The abundance of sequences homologous to the 41-bp repeats in rice support this suggested stage. (iv) Shuffling by virtue of the class II transposon. The MITE-intermediate-type TE, as suggested for the aforedescribed SCAR sequence may have evolved from a series of class II-derived movements and recombinations.

Most DNA recombinations take place during meiosis (Zickler and Klencker, 1999). However, grapevines are propagated clonally and (except in the course of breeding programs) do not undergo sexual reproduction. In the absence of a meiotic phase, the chances for DNA recombination are minimal, and the original insertion is more likely to be conserved. However, the aforedescribed data indicate that few recombination events did occur. Therefore, the PVY sequence was inserted into a one, or a few plants of a grapevine progenitor prior to cultivation, and further rearrangements took place then, or during a breeding program. This may explain why only some grapevine varieties carry the potyviral sequence.

Genome integration of sequences originated from other viral RNAs was recently reported (Crochu et al., 2004). Flaviviral-homologous sequences were found to occur in the mosquito genome. Potyviruses and flaviviruses share a common replication pathway: A positive strand of a viral RNA is translated into a polyprotein that is then processed to produce mature viral proteins. In the case of the flaviviruses, about two thirds of the viral genome was found integrated into the mosquito genome. In one case, a single ORF of 1557 amino acids was observed, but most of the flavi-like sequences are small and dispersed, indicating a series of recombination events. No indication of transposition was reported for the flaviviruses' integration case.

As already noted, recombination between viral RNAs and transcripts of viral transgenes have been previously reported. The potyviral insertion may be regarded as a "native transgene" and may play a role in the development of diversity among potyviruses.

RNA recombination was reported to occur with viral RNAs of the same, or related viruses, likely via involvement of the viral replicase (Nagy and Simon, 1997). However, non-replicative viral RNA recombination was demonstrated (Gallei et al., 2004) indicating that joining of RNA pieces may occur in the absence of the viral replicase. Processing of ribosomal and transfer RNAs, as well as splicing of primary transcripts, are in fact, a form of RNA recombination. Indeed, cases of recombination between viral and host RNAs have been reported (Charini et al., 1994; Meyers et al., 1989; Monroe and Schlesinger, 1983). This paper presents evidence of possible recombination between a viral RNA and a retrotransposable element with the product subsequently inserted into the host cell genome.

An alternative hypothesis is that the potyviral sequences originated from a plant genome. At this point, we do not favor this theory because potyviral sequences were found in only some grapevine varieties and not others. Hence, they are unlikely to be a native component of the grapevine genome. Potyviruses are widespread and polyphagous. If the phenomenon is of a general nature, then horizontal carryover of a sequence from one plant species to another could account partially for this property.

As already speculated by Crochu et al. (2004), these findings may represent an entirely new avenue of generating diversity in Eukaryotic cells.

Materials and methods

Initial detection of potyviral sequences in the plants' genome and detection of potyviral protein

Plant DNA (grapevine and tobacco) was extracted according to Bernatzky and Tanksley (1986). PCR for detection of potyviral sequences in plant DNA was performed with primers 1 and 2 for grapevine and 3 and 4 for tobacco (Table 2). Expression of a potyviral protein was assayed by ELISA (Clark and Adams, 1977) with rabbit antibodies prepared in house (courtesy D. Heller) from purified PVY.

Southern, Western, and RNase protection assays

Southern analysis was carried out according to routine procedures (Sambrook and Russel, 2001). DNA was extracted from grapevine , treated with RNase A, cleaved with *Eco*RI, and electrophoresed. Following blotting, the membrane was reacted with a cloned ^{32}P-labeled fragment of PVY-CP.

To detect transcription of PVY sequences we employed RNase protection assays. RNA was extracted with the RNeasy Plant Mini Kit (Qiagen) and hybridized with a ^{32}P-labeled in vitro transcribed fragment of a PVY-CP clone (antisense orientation, ca. 400-base-long) as described by Zeitoune et al. (1999).

Western blot analyses were carried out according to Towbin et al. (1979). The membrane was reacted with anti-PVY antibodies (prepared in our laboratory) and the relevant bands were detected by chemiluminescence employing the Supersignal West Pico Kit (Pierce).

Genome walking

Genome walking from the potyviral sequence into the flanking grapevine sequences was carried out using Clonetech's Universal genome Walker Kit and according to the manufacturer's protocol. The DNA for the genome walking experiments was RNase-treated prior to cleavage with restriction enzymes. All genome-walking primers were 29–33 bases long with *T*m values about 70 °C. Primers used for genome walking are listed in Table 2. Primers 5 and 7 were designated for extension of the PVY sequences downstream to the 3′ UTR of PVY and upstream to the PVY coat protein gene, respectively. Primers 6 and 8 served as nested primers to the above. The resultant genome walking PCR products were cloned into the vector pDrive (Qiagen) and sequenced therefrom. PCR products carrying the sequences of primer nos. 6 or 8 were selected for analysis of adjacent plant sequences flanking the PVY segment. Primers 9 and 10 were designated according to the sequence of the grapevine SCAR marker found to flank the potyviral sequence in clone GWA5. Genome walking, in this case, started at the grapevine sequence and went across the potyviral sequence to the other end of GWA5.

References

Adams, S.D., Tzeng, W.P., Chen, M.H., Frey, T.K., 2003. Analysis of intermolecular RNA–RNA recombination by rubella virus. Virology 309, 258–271.

Ashby, M.K., Warry, A., Bejarano, E.R., Khashoggi, A., Burrell, M., Lichtenstein, C.P., 1997. Analysis of multiple copies of geminiviral DNA in the genome of four closely related Nicotiana species suggest a unique integration event. Plant Mol. Biol. 35, 313–321.

Bar-Joseph, M., Yang, G., Gafny, R., Mawassi, M., 1997. Subgenomic RNAs: the possible building blocks for modular recombination of closterovirus genomes. Semin. Virol. 8, 113–119.

Bejarano, E.R., Khashoggi, A., Witty, M., Lichtenstein, C., 1996. Integration of multiple repeats of geminiviral DNA into the nuclear genome of tobacco during evolution. Proc. Natl. Acad. Sci. U.S.A. 93, 759–764.

Bernatzky, R., Tanksley, S.D., 1986. Genetics of actin-related sequences in tomato. Theor. Appl. Genet. 72, 314–321.

Charini, W.A., Todd, S., Gutman, G.A., Semler, B.L., 1994. Transduction of a human RNA sequence by poliovirus. J. Virol. 68, 6547–6552.

Cheng, C.P., Nagy, P.D., 2003. Mechanism of RNA recombination in carom- and tombusviruse: evidence for template switching by the RNA-dependent RNA polymerase in vitro. J. Virol. 77, 12033–12047.

Clark, M.F., Adams, A.N., 1977. Characteristics of the microplate method of enzyme-linked immunosorbent assay for the detection of plant viruses (ELISA). J. Gen. Virol. 34, 475–483.

Crochu, S., Cook, S., Attoui, H., Charrel, R.N., de Chesse, R., Belhouchet, M., Lemasson, J.-J., de Micco, P., de Lamballerie, X., 2004. Sequences of flavivirus-related RNA viruses persist in DNA form integrated in the genome of *Aedes* spp. mosquitoes. J. Gen. Virol. 85, 1971–1980.

Dzianott, A., Bujarski, J.J., 2004. Infection and RNA recombination of Brome mosaic virus in *Arabidopsis thaliana*. Virology 318, 482–492.

Fechotte, C., Mounches, C., 2000. Evidence that a family of miniature inverted-repeat transposable elements (MITEs) form the *Arabidopsis thaliana* genome has arisen from a pongo-like DNA transposon. Mol. Biol. Evol. 17, 730–737.

Flint, S.J., Enquist, L.W., Racaniello, V.R., Skalka, A.M., 2004. Principles of Virology, 2nd ed., ASM Press, Washington, DC (Chapter 7: Reverse Transcription and Integration).

Gallei, A., Pankraz, A., Thiel, H.-J., Becher, P., 2004. RNA recombination in vivo in the absence of viral replication. J. Virol. 78, 6271–6281.

Goff, S.P., 1992. Genetics of retroviral integration. Annu. Rev. Genet. 26, 527–544.

Harper, G., Osuji, J.O., Heslop-Harrison, J.S., Hull, R., 1999. Integration of banana streakbadnavirus into the Musa genome: molecular and cytogenetic evidence. Virology 255, 207–213.

Hu, W.S., Temin, H.M., 1990. Retroviral recombination and reverse transcription. Science 250, 1227–1233.

Jakowitsch, J., Mette, M.F., van der Winden, J., Matzke, M.A., Matzke, A.J.M., 1999. Integrated pararetroviral sequences define a unique class of dispersed repetitive DNA in plants. Proc. Natl. Acad. Sci. U.S.A. 96, 13241–13246.

Lazarowitz, S.G., 2001. Plant viruses. In: Knipe, D.M., Howly, P.M. (Eds.), Fields Virol, vol. 1. Lippincott Williams and Wilkins, Philadelphia, PA, USA, pp. 559–598.

Le, Q.H., Wright, S., Yu, Z., Bureau, T., 2000. Transposon diversity in *Arabidopsis thaliana*. Proc. Natl. Acad. Sci. U.S.A. 97, 7376–7381.

Lockhart, B.E., Menke, J., Dahal, G., Olszewski, N.E., 2000. Characterization and genomic analysis of tobacco vein clearing virus, a plant pararetrovirus that is transmitted vertically and related to sequences integrated in the host genome. J. Gen. Virol. 81, 1579–1585.

Mette, M.F., Kanno, T., Aufsatz, W., Jakowitsch, J., van der Winden, J., Matzke, M.A., Matzke, A.J.M., 2002. Endogenous viral sequences and their potential contribution to heritable virus resistance in plants. EMBO J. 21, 461–469.

Meyers, G., Rumenapf, T., Thiel, H.-J., 1989. Ubiquitin in a togavirus. Nature 341, 491.

Monroe, S.S., Schlesinger, S., 1983. RNAs from two independently

isolated defective interfering particles of Sindbis virus contain a cellular tRNA sequence at their 5′ end. Proc. Natl. Acad. Sci. U.S.A. 80, 3279–3283.

Nagy, P.D., Simon, A.E., 1997. New insights into the mechanisms of RNA recombination. Virology 235, 1–9.

Naveh, L., 1987. Detection of viruses associated with leafroll disease in grapevine by ELISA and molecular hybridization techniques. M.Sc. Thesis. The Hebrew University of Jerusalem. Faculty of Agriculture.

Ndowora, T., Dahlal, G., LaFleur, D., Harper, G., Hull, R., Olszewski, N.E., Lockhart, B., 1999. Evidence that badnavirus infection in Musa can originate from integrated pararetroviral sequences. Virology 255, 214–220.

Neuvéglise, C., Feldmann, H., Bon, E., Gaillardin, C., Casaregola, S., 2002. Genomic evolution of the long terminal repeat retrotransposons in Hemiascomycetous yeasts. Genome Res. 12, 930–943.

Oberste, M.S., Penaranda, S., Pallansch, M.A., 2004. RNA recombination plays a major role in genome change during circulation of coxsacki B viruses. J. Virol. 78, 2948–2955.

Richert-Poggeler, K.R., Noreen, F., Schwartzacher, T., Harper, G., Hohn, T., 2003. Induction of infectious petunia vein clearing (pararetro) virus from endogenous provirus in petunia. EMBO J. 22, 4836–4845.

Sambrook, J., Russel, D.W., 2001. Molecular Cloning: A Laboratory Manual. 3rd ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

Santiago, N., Herráiz, C., Goñi, R.J., Messeguer, X., Casacuberta, J.M., 2002. Mol. Biol. Evol. 19, 2285–2293.

Shapka, N., Nagy, P.D., 2004. The AU-rich RNA recombination hot spot sequence of brome mosaic virus is functional in tombusviruses: implications for the mechanism of RNA recombination. J. Virol. 78, 2283–2300.

Tanne, E., Naveh, L., Sela, I., 1989. Serological and molecular evidence for the complexity of the leafroll disease of grapevine. Plant Pathol. 38, 183–189.

Towbin, H., Staehlen, T., Gordon, J., 1979. Electrophoretic transfer of proteins from polyacrylamide gels to nitrocellulose sheets: procedure and some applications. Proc. Natl. Acad. Sci. U.S.A. 76, 4350–4354.

Wessler, S.R., Bureau, T.E., White, S.E., 1995. LTR-retrotransposones and MITEs: important players in evolution of plant genomes. Curr. Opin. Genet. Dev. 5, 814–821.

Wu, H.Y., Guy, J.S., Yoo, D., Vlasak, R., Urbach, E., Brian, D.A., 2003. Common RNA replication signals exist among group 2 coronaviruses: evidence for in vivo recombination between animal and human coronavirus molecules. Virology 315, 174–183.

Yang, G., Dong, J., Chandrasekharan, M.B., Hall, T.C., 2001. Kiddo, anew transposable element family closely associated with rice genes. Mol. Genet. Genomics 266, 417–422.

Zeitoune, S., Livneh, O., Kuznetzova, L., Stram, Y., Sela, I., 1999. T7 RNA polymerase drives transcription of a reporter gene from T7 promoter, but engenders post-transcriptional silencing of expression. Plant Sci. 141, 59–65.

Zhang, X., Feschotte, C., Zhang, Q., Jiang, N., Eggleston, W.B., Wessler, S., 2001. *P* instability factor: an active maize transposon system associated with the amplification of Tourist-like MITES and a new superfamily of transposases. Proc. Natl. Acad. Sci. U.S.A. 98, 12572–12577.

Zickler, D., Klencker, N., 1999. Meiotic chromosomes: integrating, structure and function. Annu. Rev. Genet. 33, 603–745.