Theor Appl Genet (2002) 105:614–621
DOI 10.1007/s00122-002-0969-2

F. Pelsy · D. Merdinoglu

Complete sequence of *Tvv1*, a family of Ty*1 copia*-like retrotransposons of *Vitis vinifera* L., reconstituted by chromosome walking

Received: 3 August 2001 / Accepted: 14 December 2001 / Published online: 21 June 2002
© Springer-Verlag 2002

Abstract A chromosome-walking strategy was used to sequence and characterize retrotransposons in the grapevine genome. The reconstitution of a family of retroelements, named *Tvv1*, was achieved by six successive steps. These elements share a single, highly conserved open reading frame 4,153 nucleotides-long, putatively encoding the *gag*, *pro*, *int*, *rt* and *rh* proteins. Comparison of the *Tvv1* open reading frame coding potential with those of drosophila *copia* and tobacco *Tnt1*, revealed that *Tvv1* is closely related to Ty*1 copia*-like retrotransposons. A highly variable untranslated leader region, upstream of the open reading frame, allowed us to differentiate *Tvv1* variants, which represent a family of at least 28 copies, in varying sizes. This internal region is flanked by two long terminal repeats in direct orientation, sized between 149 and 157 bp. Among elements theoretically sized from 4,970 to 5,550 bp, we describe the full-length sequence of a reference element *Tvv1-1*, 5,343 nucleotides-long. The full-length sequence of *Tvv1-1* compared to pea *PDR1* shows a 53.3% identity. In addition, both elements contain long terminal repeats of nearly the same size in which the U5 region could be entirely absent. Therefore, we assume that *Tvv1* and *PDR1* could constitute a particular class of short LTRs retroelements.

Keywords Retrotransposon · Grapevine · *Vitis vinifera* L.

Introduction

Retrotransposons are closely related to retrovirus in their structure and life cycle (Boeke and Corces 1989). Both of the two major classes of retrotransposons are flanked by Long Terminal Repeats (LTRs): the Ty*1-copia* (Flavell et al. 1992; Voytas et al. 1992) and the Ty*3-gypsy* (Suoniemi et al. 1998) are ubiquitous in the plant kingdom groups, but differ from each other in their degree of sequence conservation as well as in the order of *pol* genes. The copy and paste cycle of these elements in the host genome comprise the synthesis of self-produced proteins. They are required to synthesize an extrachromosomal DNA daughter copy, by reverse transcription of an intermediate RNA, prior to its reinsertion into the genome. For a review of the mechanism of retrotransposition, see Kumar and Bennetzen (1999).

All Ty*1 copia*-like elements present the same general structure: they encode a number of proteins specified by two major genes, *gag* and *pol*. These genes and proteins are all specified by a single mRNA molecule that has the structure 5′-RU5-PBS-coding region-PPT-U3-R-3′. Boundaries between U3, R, and U5 are a consequence of the location of the promoter and polyadenylation signal in the LTR. PBS (primer binding site) and PPT (polypurine tract) are involved in the replication cycle of the element. The *gag* gene encodes proteins that assemble into virus-like particles, while *pol* genes (*pro*, *int*, *rt*, *rh*) encode respectively, protease, integrase, reverse transcriptase, and RNase H. From the mRNA molecule, a polyprotein is synthesized and then cleaved into functional peptides by the *pol*-encoded protease. Reverse transcriptase and RNase H activities are required for replication of the retrotransposon whereas, the integrase allows the DNA daughter copy to insert at a new host site, with a 5-bp host sequence duplication flanking the element. These *gag* and *pol* proteins, expressed both in retrovirus and retrotransposososns, share invariant amino-acids or highly conserved sequence stretches.

Full size Ty*1 copia*-like retrotransposons range from 3,925 bp for *PDR1* from pea (Lee et al. 1990) to 12,088 bp for *BARE-1* from barley (Manninen and Schulman 1993) whose LTRs respectively are 156-bp to 1,929-bp long. Full-length or parts of Ty*1 copia*-like retrotransposon have been identified; however, different strategies have been used for them. Some of them were

Communicated by G. Wenzel

F. Pelsy (✉) · D. Merdinoglu
Laboratoire de Génétique et Amélioration des Plantes,
UMR Vigne et Vins d'Alsace, INRA-Colmar,
28 rue de Herrlisheim, 68000 Colmar. France
e-mail: pelsy@colmar.inra.fr
Tel.: 0-(33)-3-89-22-49-00, Fax: 0-(33)-3-89-22-49-33

trapped into the coding sequence, such as *Tnt1* in the tobacco nitrate reductase gene (Grandbastien et al. 1989), *Tst1* in the potato starch phosphorylase gene (Camirand et al. 1990), or *Vine-1* in a grapevine *Adh*r gene (Verries et al. 2000). These retrotransposons can either have the complete structure, two LTRs bordering an internal domain which contains one (*Tnt1*) or more (*Tst1*) open reading frames (ORFs) that encode the *gag* and *pol* proteins, or they can be truncated (*Vine-1*). Genomic library screening has resulted in the isolation of new retrotransposons, such as *BARE-1* from barley by the use of a LTR probe (Manninen and Schulman 1993) or *PDR1* recognized as a dispersed repeated sequence in the pea genome (Lee et al. 1990). Induced by tissue culture, new retrotransposon families *Tto1–Tto3* in tobacco (Hirochika 1993) and *Tos17–Tos20* in rice (Hirochika et al. 1996) were identified by the reverse transcription-PCR method. Other retroelement sequences have emerged, like the *Basho* superfamily, from computer-based sequence similarity searches in genomic databases (Le et al. 2000). Transcriptional activity has been reported for only a few elements, mostly those characterized by trapping, such as *Tnt1* (Grandbastien 1998), or induced by tissue culture, such as *Tos17* (Hirochika et al. 1996).

Throughout a diverse range of organisms, retrotransposon copy number appears to be correlated with genome size. In the large genome of maize, retrotransposons can account for more than 50% of the nuclear DNA content, making them major constituents of this plant genome (SanMiguel et al. 1996). However, in the small Arabidopsis genome, they only compose approximately 2% of the genomic sequences (Le et al. 2000). When active these elements play a role in the genome evolution; the replicative spreading of retrotransposons in the host genome increases their copy number and creates insertional polymorphism. Stress and environmental factors can activate transposition (Grandbastien 1998), such as tissue culture (Grandbastien et al. 1989), as well as sharp climatic conditions, which increases the copy number of *BARE-1* in wild barley (Kalendar et al. 2000).

Grapevine is one of the most important fruit crops of the world. The genome size is approximately 475 M bp/1C, 96% of which is most likely repetitive or non-coding in nature (Lodhi and Reich 1995). Previous work has shown the presence of retrotransposons: Verries et al. (2000) have identified a truncated Ty*1 copia*-like retrotransposon inserted in an *Adh*r gene and, in addition, Böhm and Zyprian (1998) have generated two partial retrotransposon sequences by RADP (690-bp long). In this study, we describe the step by step reconstitution by chromosome walking of a family of grapevine retrotransposons named *Tvv1*. These elements show a full-length ORF, this study being the first to describe such a grapevine retrotransposon, whose order of *pol* genes is typical of Ty*1 copia*-like retrotransposons.

Materials and methods

Plant DNA

This work was performed on *Vitis vinifera* L. riesling cultivar 49 (ENTAV et al. 1995). Young, expanding leaves of an individual plant were collected in greenhouses then ground into a fine powder with liquid nitrogen. Total DNA was extracted according to the Lodhi et al. (1994) protocol with slight modifications: chloroform:isoamyl alcohol (24:1 v/v) was used instead of chloroform:octanol (24:1 v/v), then a RNase treatment was performed before DNA precipitation.

Generation and cloning of PCR fragments

Five independent pools of grapevine DNA were created by digestion with five restriction enzymes, then a unique adaptator sequence was ligated to the digested fragments. To generate each step, the chromosome-walking method consisted of two successive PCR amplifications. The first one amplified fragments between a primary gene-specific primer and the outer adaptator primer; this primary amplification product was then used as a template for a second amplification between a nested gene-specific primer, which did not overlap the primary primer, and a nested adaptator primer. We designed each pair (primary and nested) of retrotransposon-specific primers, 27-mers long, from the previous sequenced step and synthesized them by MWG Biotech AG (Ebersberg, Germany). The sequence and origin of primers is given in Table 1.

The Universal GenomeWalker Kit (Clontech, Palo Alto, Calif.) was used to generate the retrotransposon-specific PCR fragments. According to the supplier's protocol, five aliquots of total DNA were separately hydrolyzed with one of the five restriction enzymes of the kit, *Eco*RV, *Sca*I, *Dra*I, *Pvu*II and *Stu*I. Then each pool of digested DNA fragments was ligated to the GenomeWalker adaptator. Amplification reactions were performed in a 20-µl reaction mixture with the 50× Advantage Genomic PCR Kit (Clontech, Palo Alto, Calif.), in a Touch Down Thermal Cycling System (Hybaid, Middlesex, UK), according to Clontech instructions. The first PCR program consisted of seven cycles of 94 °C for 10 s, and 70 °C for 3 min, followed by 32 cycles of 94 °C for 10 s, and 65 °C for 3 min. The second program consisted of five cycles of 94 °C for 10 s, and 70 °C for 3 min, followed by 20 cycles of 94 °C for 10 s, and 65 °C for 3 min. Finally both programs were followed by an elongation step of 65 °C for 10 min.

Aliquots of the PCR products were analyzed by electrophoresis on a 1.5% agarose gel in 1 × TAE (40 mM TRIS-acetate pH 7.8, 20 mM EDTA).

For a band of interest, at least three independent PCR fragments cloned into the pCR 4-TOPO cloning vector (InVitrogen, Groningen, The Netherlands), according to the supplier's instructions, were selected by size and sequenced.

PCR amplification

PCR was carried out in the GeneAmp PCR System 9700 thermocycler (Perkin Elmer Biosystems, Foster City, Calif.), and the *Taq* DNA polymerase (Amersham Pharmacia Biotech Inc, Piscataway, N.J.). When primers Pltr1 IRD 800 5′ labeled-P17 were used in couple, the cycling program consisted of the following: 94 °C for 10 min, 30 cycles of 92 °C for 1 min, 52 °C for 1 min and 72 °C for 2 min, followed by 72 °C for 4 min. Amplified fragments were resolved by electrophoresis on a 60-cm long acrylamide gel in a LiCor 4000 L automated DNA sequencer (Lincoln, NB) with IRD41-labeled M13 fragments used as a standard of the molecular weight. When primer P18 and Pltr4 were used, PCR conditions were as follows: 94 °C for 2 min, 35 cycles of 94 °C for 10 s, 55 °C for 30 s and 68 °C for 8 min, followed by 68 °C for 7 min. Amplification products were resolved by electrophoresis on a 1% agarose gel.

Table 1 Primers used. GW primers have been employed in the chromosome walking procedure

Primer	Orientation	Sequence (5′–3′)	Originated from fragment	Position/*Tvv1-1*	Step generated
GW1	Sense	CATTTGGCATGAACAAGAGCGAGAAAG	RTD121	4,150–4,176	
GW2	Sense	CTGTTGGTATCATACTTCTTGTGGTCT	RTD121	4,204–4,230	B
GW5	Antisense	TACTTCAGTTCACCCAAGTCCTTTGTA	5′-consensus B	4,310–4,336	
GW6	Antisense	GACATAGACCACAAGAAGTATGATACC	5′-consensus B	4,209–4,235	N
GW7	Antisense	GCGAGAAACATCTATAGCCCTTTTGTA	5′-consensus N	3,178–3,204	
GW8	Antisense	TTGAGTACCCCAAGAAAACACACTGTA	5′-consensus N	3,148–3,174	P
GW9	Antisense	CCAAGTCATTCGAGAAAAATCATCCAC	5′-consensus P	2,649–2,675	
GW10	Antisense	CCAGTTTGAGAAGTAACAGGACACGGA	5′-consensus P	2,600–2,626	R
GW13	Antisense	GGACGATGAAAGGCATTCCATACATCA	5′-consensus R	1,373–1,399	
GW14	Antisense	ACATCCGAGATACATTTCCTTTTCCAG	5′-consensus R	1,345–1,371	T
GW15	Antisense	TCTGCAGAAATAACCGTGCATCATCTT	5′-Consensus T	1,228–1,254	
GW16	Antisense	GTGGTCATCTTTAGCAACACTTCTCAA	5′-Consensus T	1,155–1,181	Z
P17	Antisense	(CT)AGAATTCTTACTCTCTTCC	5′ ORF	1,043–1,062	
Pltr1	Sense	CCTAATTCAGGACTCTCAAT	LTR1	61–80	
P18	Sense	GTTGCTGATATAGTGCCTATTGTGTCA	5′ ORF	1,062–1,088	
Pltr4	Antisense	CAATTGAGAGTCCT(AG)AATTAGGAAGGT	LTR1	5,250–5,278	

DNA sequence determination and analysis

Sequencing was done by the dideoxy method, using the SequiTherm II Long-read DNA Sequencing Kit (Epicentre Biotechnologies, Madison, Wis.) and separated on a 60-cm long acrylamide gel by a LiCor 4000 L automated DNA sequencer.

Computer-assisted analysis of the sequence data was performed using programs of the Wisconsin Sequence Analysis Package (version 9.0). Alignments of the hypothetical translation of *Tvv1* to other elements was made from the following sequence accessions: *copia* (*Drosophila melanogaster* retrotransposon, X02599), *Tnt1* (*Nicotiana tabacum* retrotransposon, X13777), *PDR1* (*Pisum sativum* retrotransposon, X66399) and *Vine-1* (*V. vinifera* retrotransposon, AF116598). The nucleotide sequence data of the 27 reported fragments will appear in GenBank database under accession numbers: AF478364 to AF478390.

Results

Walking along the chromosome

The first pair (primary and nested) of retrotransposon specific primers – sense – were defined from a 250-bp long sequence (RTD121) obtained by the amplification of grapevine DNA between two degenerated primers corresponding with conserved reverse transcriptase domains I and III of Ty*1 copia*-like retrotransposons (This, personal communication), enabling the generation of the B step. For the following steps, we designed the pair of antisense retrotransposon-specific primers from the previous sequence, close to its 5′ end, in order to generate overlapping sequences to establish a contig sequence. Primary and nested amplifications of the five DNA pools were performed at each step, creating a single amplified product from only one or two of the pools, except for steps B and Z which showed multiple-band profiles. From each amplified product, three to five independent PCR fragments were cloned and sequenced.

One step downstream (B) and five successive steps upstream (N, P, R, T, Z) from the initial 250-bp sequence generated fragments ranging from 395 to 1,590 bp, with overlapping sequences ranging from 45 to 319-bp further used to establish a contig sequence of a maximum size of 6,047 bp (Fig. 1).

Table 2 Sequence identity between fragments taken from the same step and located in the *Tvv1* ORF. The reference fragment is indicated in bold. Only the highly conserved part of the B and T fragments have been taken into account, indicated by a star

Step	Fragment	Size (bp)	Sequence identity (%)
B*	4b11	990	
	4b2	990	98.86
	1b5	991	96.43
N	5n7	1,119	
	5n5	1,119	96.60
	3g1	677	95.57
P	2p5	593	
	2p6	593	99.32
	2p8	593	97.80
R	1r3	1,349	
	1r1	1,348	97.32
	1r4	1,353	98.22
	1r6	438	98.17
T*	2t7	394	
	2t9	394	97.97
	3t2	351	98.57

The identity level of the different fragments taken from the same step have been evaluated. Full-size fragments of steps N, P and R, and part of the fragment sequences of steps B and T, show an identity level higher than 95% (Table 2). Stretches of sequences, around 150 bp, of steps B and Z show an identity level ranging between 98 and 70%, while parts of steps B, T and Z are much more variable. Based on these identity levels, respectively, three classes of sequences have been determined (Fig. 1).

In the three classes, the nucleotide sequence of each cloned fragment differed from the others by more than 2 bp in every 100 bp, which is a higher frequency than

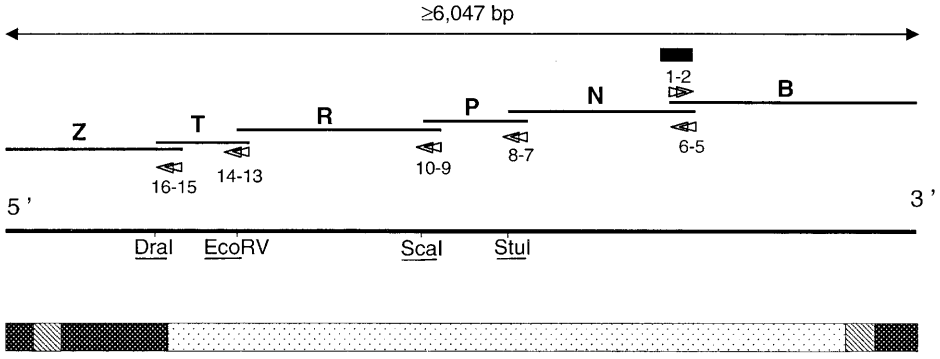

Fig. 1 Schematic representation of the chromosome-walking procedure. The *black box* corresponds to the initial 250 bp-long sequence. *Double arrows* represent the pairs of primary and nested primers used to generate each step indicated by *capital letters*. *Thick lines* indicate the contig sequence with the position of the different restriction sites corresponding to the enzymes used to create the DNA pools. Below; the three classes based on the identity level of the PCR fragments taken from the same step are represented: ⬚ Class 1: identity higher than 95%. ▨ Class 2: identity ranging between 98 and 70%. ▩ Class 3: variable sequences

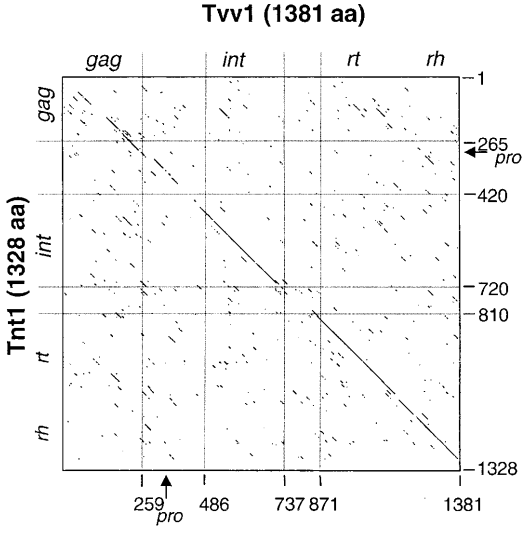

Fig. 2 Homology matrix comparisons between the amino-acid sequence of the *Tvv1* and the *Tnt1* ORFs. The approximate position of the regions corresponding to the *gag* protein and to the *int*, *rt* and *rh* domains encoded by *pol* genes are indicated between *dotted lines*. The active site of the protease *prot* is shown by an *arrow*. A match of 11 amino-acid residues within a window of 30 produces a *dot*

that for base substitution (0.1%/25 cycles) induced by the polymerase mix. This result supports the hypothesis that these fragments did not derive from a unique retrotransposon copy, but rather from related copies present in the grapevine genome, sharing the same gene-specific primer sites.

Open reading frame in *Tvv1*

From fragment sequences of class 1, a consensus sequence has been deduced at each step and then a contig sequence 4,153-bp long. Hypothetical translation of this contig sequence revealed a single putative ORF corresponding to a 1,381 amino-acid-long polyprotein. Any stop-codon interrupts this putative ORF; however, when we analyzed the hypothetical amino-acid sequences of each of the actual 17 fragment sequences (from steps T, R, P, N and B) located in this region, two exhibit a stop-codon and three a frameshift, a consequence of the addition of one nucleotide for two of them and of a 32 bp-long duplication for the third.

By comparison with the *copia* and *Tnt1* ORFs, the general organization of this putative ORF is very similar both in structure, as it encodes the *gag*, *pro*, *int*, *rt* and *rh* proteins in this order, and in length, 1,328 amino-acids for *Tnt1* and 1,404 amino-acids for *copia*. This ORF appears to belong to a Ty*1 copia*-like retrotransposon named *Tvv1* (Transposon *V. vinifera*). A homology matrix comparison (Fig. 2) between the amino-acid sequences of the *Tvv1* and *Tnt1* ORFs shows different levels of homology depending on the domain, integrase and reverse transcriptase-RNaseH are the most conserved domains with successively 42% identity on 252 amino-acids and 40.2% identity on 517 amino-acids, although, the GAG polyprotein is less conserved with 28.7% identity on 105 residues. In addition, several short stretches of amino-acids, shown to be highly conserved or invariant in all retrotransposons, have been found in the *Tvv1* ORF (Fig. 3). Small differences in conserved stretches of protease and RNaseH motif II show that *Tvv1* could be more closely related to *copia* than to *Tnt1*. This contig sequence is ended by the sequence TGAGGGGGAG corresponding to the consensus PPT.

Tvv1 LTR

Class 2 corresponds to sequences ranging between 151 and 157 bp, observed in five Z fragments, and between 149 and 153 bp in nine B fragments. Moreover, the identity level of these 14 sequences is at least 70%. They thus correspond to repeats in direct orientation. All of them show the consensus motifs 5′-TG…AC-3′ expected

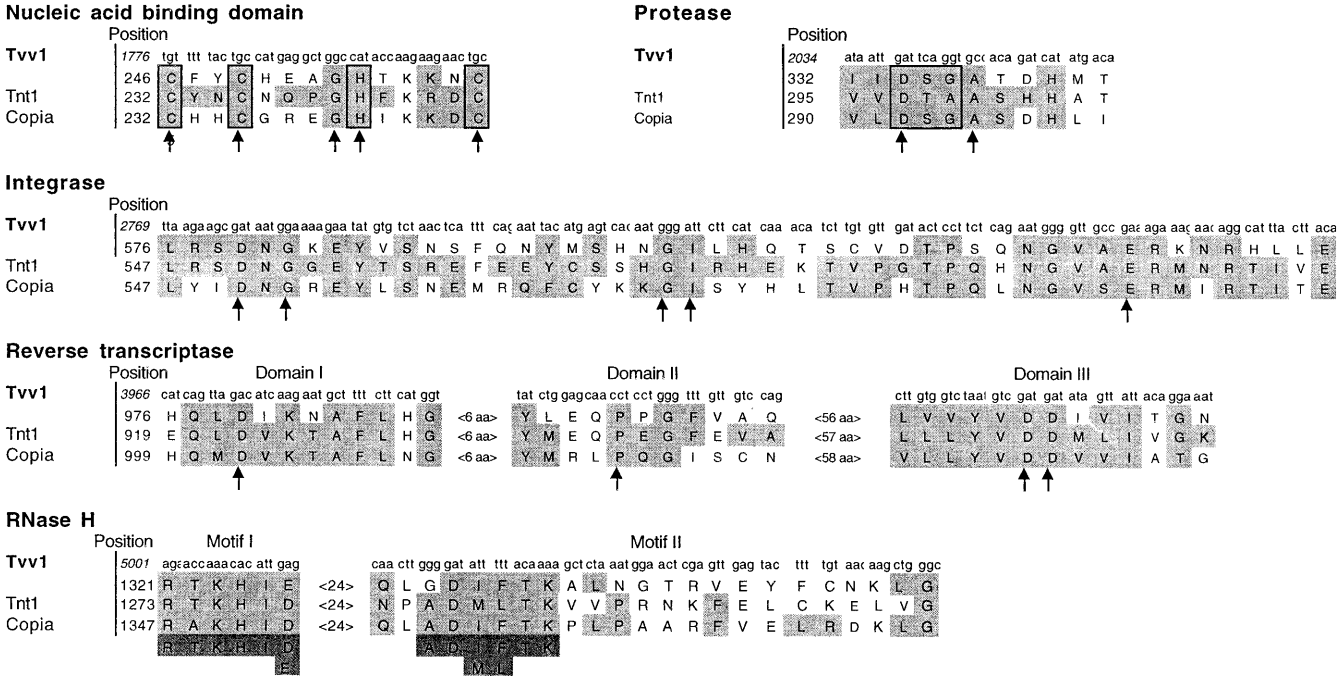

Fig. 3 Comparison of the hypothetical *Tvv1* ORF with conserved domains found in *Tnt1* and *copia*. Amino-acids are indicated by *capital letters*, and the corresponding nucleotide sequence by *lower-case letters*. Conserved or nearly invariant amino-acids are shared. Fully conserved positions are indicated by *arrows*. In the reverse transcriptase and RNase H domains, the number of amino-acids separating the motifs are indicated between *angle brackets*. *Boxes* in the nucleic acid-binding domain show the cysteine and histidine residues involved in the zinc-binding domain, and the *box* in the protease domain shows the conserved D-T-(S)-G amino-acid sequence box. *Numbers* indicate the position of the amino-acids in the ORFs and of the nucleotides in *Tvv1-1*

Fig. 5 Alignment of three consensus LTR sequences. The putative TATA box and polyadenylation signal are *underlined*. Nucleotides different from LTR1 are in *grey*

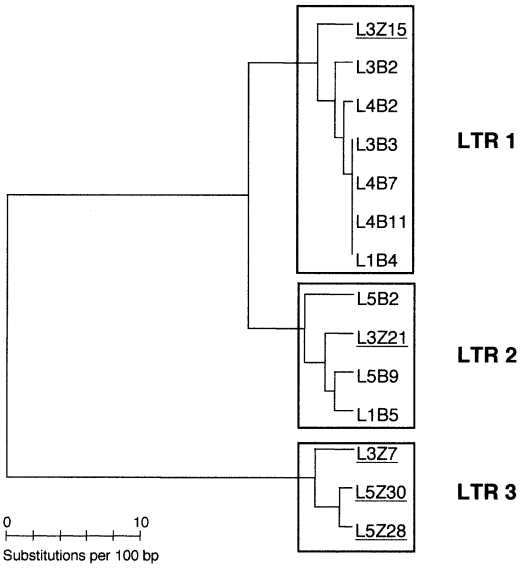

Fig. 4 Dendrogram resulting from the UPGMA clustering of the genetic similarity matrix of the 14 LTRs sequenced. The 5′LTRs are underlined

for long terminal repeat (LTR) sequences of retrotransposons. From Z fragments they correspond to the 5′ LTR and from B fragments to the 3′ LTR of *Tvv1*. Nei and Li's coefficient (Nei and Li 1979) genetic similarity matrix was generated with sequence data and a dendrogram was calculated by UPGMA clustering. This dendrogram reveals 11 different LTRs, suggesting that at least 11 different copies of *Tvv1* have been used as a template. It is clearly subdivided into three clusters, each of them sharing at least 95% identity (Fig. 4). Two clusters include sequences from the Z and B step. From each cluster a consensus sequence has been deduced: LTR 1 (150-bp long), LTR 2 (151 bp) and LTR 3 (156 bp) (Fig. 5).

Three discrete regions, U3, R and U5, can be found in typical retroviral or retrotransposon LTRs (Temin 1981). The sequence ATAATA, similar to the consensus TATA box (Joshi 1987) present in the three LTRs (position 51–56), is the potential TATA box usually located in the

U3 region (Fig. 5). The U3/R boundary could then be the G residue, at position 83, 26 bp after the TATA box, in accordance with the 23-bp distance found in the retrovirus LTR (Temin 1981). The sequence AATAAA similar to the consensus polyadenylation signal (Mogen et al. 1990) at position 113–118, is usually in the R region. No CA motif that marks the R/U5 boundary can be found in the 40 nucleotides behind the above identified polyadenylation signal. Therefore, no R/U5 boundary seems to exist supporting the conclusion that U5 is entirely absent in these *Tvv1* LTRs, as has been previously observed for *PDR1*.

Regions located between the 5′ LTR and the ORF

Class 3 of the variable sequences corresponds to: (1) host sequences in front of the 5′ LTR and behind the 3′ LTR, and (2) an internal region located between the 5′ LTR and the ORF (Fig. 1).

The host sequences in five Z and nine B fragments show a high variability in length (from 15 to 452 bp) and in sequence. In addition, host 5-bp stretches immediately flanking the LTRs, which result in direct repeats created upon retrotransposon insertion, are also highly variable. This is in accordance with the hypothesis that the Z and B fragments containing LTRs have been amplified from at least 11 different related copies of *Tvv1* located at different sites in the grapevine genome.

In the five Z and partially in the five T fragments, regions located between the 5′ LTR and the ORF are heterogeneous. We have sequenced one 890 bp-long sequence (in the 3Z15 fragment) and four 511–515-bp sequences (in the 3Z21, 3Z7, 5Z30 and 5Z28 fragments). Immediately starting all of these sequences the identical sequence TGGTATCA (G/T) AGCC, which corresponds to the PBS complementary to the $tRNA^{met}$, has been found. The identity level of the last four sequences was 88% and a consensus sequence 513-bp long has been deduced. Comparison of the 513-bp consensus to the 890 bp-long fragment shows a 45.3% identity, mostly located in the 80 bp upstream of the ORF. Neither of these two sequences revealed a reliable ORF like those of most retrotransposons, or any clear homology with bank sequences.

To evaluate the range of size of this untranslated leader region in the different *Tvv1* elements of an individual plant, we made a PCR between primers Pltr1 located in a consensus part of 5′ LTR and P17 located at the beginning of the ORF. A set of 28 major bands was scored, approximately sized between 630 bp to 1,200 bp, which corresponded to untranslated leader regions sized between approximately 520 to 1,090 bp (Fig. 6). This result may be explained by assuming that these bands belong to *Tvv1* variants sharing the same primers sites; hence, these *Tvv1* elements most likely constitute a family theoretically sized from 4,970 to 5,550 pb.

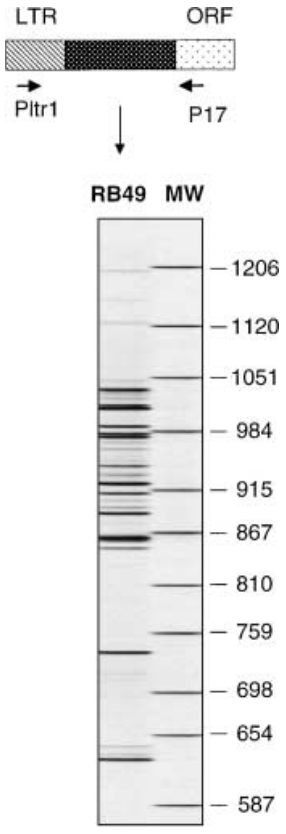

Fig. 6 Fingerprinting results obtained using the IRD800 fluorescence dye-labeled Pltr1 and P17 primers. The molecular weight has been prepared by IRD41-labeled amplifications of M13

Reconstitution of two reference *Tvv1* elements

From the consensus sequences of the different parts of elements from this *Tvv1* family, a reference element has been reconstituted named *Tvv1-1*. It has been built with the 5′ LTR 1 (150 bp), the untranslated leader region (890 bp) of fragment 3Z15 behind the sequence LTR from the LTR 1 cluster, the common ORF-PPT sequences (4,153 bp), and 3′ LTR 1(150 bp). Thus, it constitutes a 5,343-bp typical Ty*1 copia*-like retrotransposon. A variant of *Tvv1-1*, *Tvv1-2*, has been built with the 5′ LTR 2 (151 bp) from the LTR 2 cluster, the consensus untranslated leader region (513 bp), the common ORF-PPT sequences (4,153bp), and the 3′ LTR 2 (151 bp). It constitutes a 4,968-bp retrotransposon.

The full-length sequence of *Tvv1-1* has been compared to four Ty*1 copia*-like retrotransposons, shown with a pea *PDR1* 53.3% identity, a potato *Tst1* 51.7% identity, and with a *copia* 43.9% identity and a *Tnt1* 43.7% identity. But with grapevine retrotransposon *Vine-1* the identity level decreased to 38.8%, indicating that *Vine-1* and *Tvv1* are two distinct Ty*1 copia*-like retrotransposons of the same species.

To check the actual length of the *Tvv1* ORF, two primers have been designed, P 18 at the very beginning of the ORF and Pltr3 in the 3′ LTR, making possible the amplification of a 4,215 bp-long product. Amplification

created a broad band corresponding to a major product of approximately 4.3 kb, in agreement with the structural organization we previously determined (data not shown).

Discussion

By six successive steps we achieved the reconstitution of grapevine retroelements belonging to a family named *Tvv1*. These elements share a single highly conserved ORF 4,153-nucleotides long, a variable untranslated leader region, upstream of the ORF, as well as PBS and PPT sites flanked by two repeats in direct orientation, sized between 149 and 157 bp. Comparison of the consensus ORF sequence with those of *copia* and *Tnt1* revealed that *Tvv1* elements are closely related to both *copia* and *Tnt1*, and could encode for the proteins *gag*, *pro*, *int*, *rt* and *rh*, in this order, of a typical Ty*1 copia*-like retrotransposon. Based on the hypothetical translation of the consensus sequence, the amino-acid sequence of the *Tvv1* ORF is not interrupted by any stop codon and could encode a full-length putative polyprotein. It can be expected that some *Tvv1* elements could synthesise the necessary functions for their autonomous transposition by reverse transcription and be transpositionally active in the grapevine genome. However, as the hypothetical translation of a few fragments located in this coding region are stopped by a stop codon or a frameshift, corresponding *Tvv1* elements must be inactivated. The mutations leading to stop codons or frameshifts could result in the error-prone activity of the reverse transcriptase. During replication of Ty*1*, only base substitutions were observed with a rate of 2.5×10^5 bp per cycle (Gabriel et al. 1996). However, mutations occurring during retroviral replication include base substitutions and frameshifts, as well as a complex deletion or a deletion with an insertion. Upon all sequenced fragments, only one (3B2) showed a 32-bp duplication at the end of the RNase H domain, giving rise to a frameshift.

Evidence for transpositional activity of a retrotransposon can also be inferred from the analysis of its LTR sequences, which are identical in newly transposed copies. When comparing *Tvv1* LTR sequences, no 5′LTRs were shown identical to a 3′LTR. This could prevent the possible use as a template of a recently transposed *Tvv1* element in our study. It has been confirmed by the absence of the same 5-bp host duplication, generated by insertion of a retrotransposon into a new site.

The *Tvv1* LTRs are from 150 bp to 155 bp and show a minimum of 70% variability. These LTRs are shorter than *copia* and *Tnt1* LTRs, respectively 276 bp and 610 bp. Moreover, *Tvv1* LTRs seem to have no R/U5 boundary so that the U5 region could be entirely absent. Both their size and absence of the U5 region are particularities shared with the *PDR1* 156-bp LTRs (Lee et al. 1990). However, no size variation of the length of *PDR1* LTRs has been shown (Vershinin and Ellis 1999). In addition, comparison of the full-length sequences of *Tvv1* and *PDR1* show 53.3% identity; thus, these two elements could constitute a particular class of short LTRs retroelements.

If the sequence corresponding to the ORF is highly conserved on all *Tvv1* elements, the untranslated leader region is variable both in size and sequence. Amplification of that precise region shows that *V. vinifera* cv riesling contains at least 28 copies of *Tvv1*, in varying sizes. We think that *Tvv1* elements form a family of retrotransposons sharing the coding region, although major differences in the untranslated leader region that most likely result in an accumulation of modifications by recombination or mutations, could have an effect on the retrotransposition capacity of these elements.

The *PDR 1* internal structure does not contain any untranslated leader regions in front of the ORF, when such a leader sequence is present in *copia* (300 bp) (Emori et al. 1985), *Tnt1* (461 bp) (Pouteau et al. 1991) and *BARE 1* (2,057 bp) (Manninen and Schulman 1993). Heterogeneity of the internal structure of the *PDR* variants shows differences in the size of their 5′-region corresponding to the *gag* domain, although other functional domains remain highly conserved (Vershinin and Ellis 1999). The existence of subfamilies and variants that differ by large re-arrangements in the ORF have been described for the WIS-2 element in wheat (Moore et al. 1991), as well as in the untranslated leader region for *BARE-1* in barley (Shcherban and Vershinin 1997).

Verries et al. (2000) were the first to report the characterization of a LTR retrotransposon in grapevine, *Vine-1*, albeit defective. This truncated element 2,392-bp long, which is related to Ty*1 copia*-like retroelements, contains two almost identical 287-pb LTRs. In this study, we describe *Tvv1* elements which most-likely constitute a family theoretically sized from 4,970 to 5,550 bp, sharing a full-length ORF, the first described for a grapevine retrotransposon. Comparison of the full-length sequences of the reference elements *Tvv1-1* and *Vine-1*, revealed 38.8% homology. The differences in LTR size and sequence (43.3% identity with LTR 1) and the low sequence identity led us to infer that these two retrotransposon families of the same species are distinct.

This chromosome-walking method has been efficient to sequence 5-kb retroelements from a 250 bp starting point, containing conserved reverse transcriptase motifs and using at least 11 different copies of *Tvv1* as a template. Therefore, we demonstrate the efficiency of this method to characterize an even large multicopy template so that it could be applied to characterize any species retrotransposon.

Acknowledgements The authors thank Lucie Schehrer and Vincent Chiquet for valuable technical assistance, and Katherine Olson for her careful revision of the English. All experiments described in this study comply with the current laws of France. This research was partly supported by a grant of the Ministère de l'Agriculture et de la Pêche (Contrat CTPS n: B 02528).

References

Boeke JD, Corces VG (1989) Transcription and reverse transcription of retrotransposons. Annu Rev Microbiol 43:403–434

Böhm A, Zyprian E (1998) RAPD markers in grapevine (*Vitis* spp.) similar to plant retrotransposons. Plant Cell Rep 17:415–421

Camirand A, St-Pierre B, Martineau C, Brisson N (1990) Occurrence of a copia-like transposable element in one of the introns of the potato starch phosphorylase gene. Mol Gen Genet 224: 33–39

Emori Y, Shiba T, Kanaya S, Inouye S, Yuki S, Saigo K (1985) The nucleotide sequences of *copia* and *copia*-related RNA in *Drosophila* virus-like particles. Nature 315:773–776

ENTAV, INRA, ENSAM, ONIVINS (1995) Catalogue des variétés et clones de vigne cultivés en France

Flavell AJ, Dunbar E, Anderson R, Pearce SR, Hartley R, Kumar A (1992) *Ty1-copia* group retrotransposons are ubiquitous and heterogeneous in higher plants. Nucleic Acids Res 20:3639–3644

Gabriel A, Willems M, Mules EH, Boeke JD (1996) Replication infidelity during a single cycle of *Ty1* retrotransposition. Proc Natl Acad Sci USA 93:7767–7771

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

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

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

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

Joshi C (1987) An inspection of the domain between the putative TATA box and translation start site in 79 plant genes. Nucleic Acids Res 15:6643–6653

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

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

Le QH, Wright S, YU Z, Bureau T (2000) Transposon diversity in *Arabidopsis thaliana*. Proc Natl Acad Sci USA 97:7376–7381

Lee D, Ellis THN, Turner L, Hellens RP, Cleary WG (1990) A *copia*-like element in *Pisum* demonstrates the use of dispersed repeated sequences in genetic analysis. Plant Mol Biol 15:707–722

Lodhi MA, Reich BI (1995) Nuclear DNA content of *Vitis* species, cultivars, and other genera of the Vitaceae. Theor Appl Genet 90:11–16

Lodhi MA, Ye GN, Weeden NF, Reisch BI (1994) A simple and efficient method for DNA extraction from grape cultivars and *Vitis* species. Plant Mol Biol Rep 12:6–13

Manninen I, Schulman AH (1993) *BARE-I*, a *copia*-like retroelement in barley (*Hordeum vulgare*). Plant Mol Biol 22:829–864

Mogen BD, MacDonald MH, Graybosch R, Hunt AG (1990) Upstream sequences other than AAUAAA are required for efficient messenger RNA3′-end formation in plants. Plant Cell 2:1261–1272

Moore G, Lucas H, Batty N, Flavell R (1991) A family of retrotransposons and associated genomic variation in wheat. Genomics 10:461–468

Nei M, Li W (1979) Mathematical models for studying genetic variation in terms of restriction endonucleases. Proc Natl Acad Sci USA 76:5269–5273

Pouteau S, Huttner E, Grandbastien MA, Caboche M (1991) Specific expression of the tobacco *Tnt1* retrotransposon in protoplasts. EMBO J 10:1911–1918

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

Shcherban AB, Vershinin AV (1997) BARE-1D, a representative of a family of *BARE*-like elements of the barley genome. Genetica 100:231–240

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

Temin HM (1981) Structure, variation and synthesis of the retrovirus long terminal repeat. Cell 27:1–3

Verriès C, Bès C, This P, Tesnière C (2000) Cloning and characterization of *Vine-1*, a LTR-retrotransposon-like element in *Vitis vinifera* L. and other *Vitis* species. Genome 43:366–376

Vershinin AV, Ellis THN (1999) Heterogeneity of the internal structure of *PDR1*, a family of Ty1/*copia*-like retrotransposons. Mol Gen Genet 262:703–713

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