Theor Appl Genet (2008) 116:671–682
DOI 10.1007/s00122-007-0700-4

ORIGINAL PAPER

Structural variability of *Tvv1* grapevine retrotransposons can be caused by illegitimate recombination

Cédric Moisy · Sophie Blanc · Didier Merdinoglu · Frédérique Pelsy

Received: 20 June 2007 / Accepted: 13 December 2007 / Published online: 10 January 2008

Abstract Structural variability of *Tvv1*, a grapevine retrotransposon Ty1 *copia*-like family, was investigated within the grape genome and the canonical sequence of *Tvv1* determined. Then, two remarkable elements, *Tvv1*-Δ3001 and *Tvv1*-Δ3640, which had suffered large deletions 3,001 bp and 3,460 bp in length of their coding sequences were compared to the canonical copy. In both deleted elements, the deletion breakpoint was characterized by a stretch 13 bp-long in *Tvv1*-Δ3001 and 11 bp-long in *Tvv1*-Δ3640 found duplicated in the canonical copy at each bound of the deleted regions. *Tvv1*-Δ3001 and *Tvv1*-Δ3460 were both shown to be unique copies fixed at a single locus in the grapevine genome. Their presence was very variable in a set of 58 varieties and wild vines. These elements have most likely been dispersed through natural intermixing after their initial insertion whose chronology was estimated. The model that we propose to explain the structure of *Tvv1*-Δ3001 and *Tvv1*-Δ3640, implies illegitimate recombination involving template switching between two RNA molecules co-packaged in the VLP prior to the integration of the deleted daughter copy into the host genome.

Communicated by C. Gebhardt.

C. Moisy · S. Blanc · D. Merdinoglu · F. Pelsy (✉)
UMR 1131 Santé de la Vigne et Qualité du Vin, INRA,
Université Louis Pasteur de Strasbourg,
68000 Colmar, France
e-mail: pelsy@colmar.inra.fr

Introduction

Retrotransposons are mobile elements closely related to retroviruses in their structure and life cycle (Boeke and Corces 1989). Active retrotransposons transpose via the reverse transcription of a transcribed, intermediate RNA, thus increasing their copy number (Feschotte et al. 2002). The copy-and-paste cycle in the host genome comprises the synthesis of specific proteins encoded by two major genes, *gag* and *pol*. The two major classes of retrotransposons, the Ty1 *copia* (Flavell et al. 1992; Voytas et al. 1992) and the Ty3 *gypsy* (Suoniemi et al. 1998), are flanked by long terminal repeats (LTRs), but differ from each other in the order of domains inside the *pol* gene as well as in their degree of sequence conservation. The *pol* gene of Ty1 *copia*-like elements encodes from 5′ the protease, integrase, reverse transcriptase and RnaseH domains. These enzymatic functions are required to synthesize an extrachromosomal DNA daughter copy from the intermediate RNA prior to its reinsertion into the genome. The *gag* gene encodes proteins that assemble into virus-like-particles (VLP) in which the genomic RNA is packaged along with the associated enzymes and primer tRNA. Ty1 and Ty3 retrotransposons VLPs contain a dimeric RNA genome similar to what has been described for retroviruses (Feng et al. 2000; Nymark-McMahon et al. 2002). However, while in retroviruses the dimer consists of two identical plus-strand RNAs joined by non-covalent bonds between stem-loops on the monomers, the mechanism of union between monomers is yet unknown for retrotransposons, although it is conceivable that they are joined to each other through the primer $tRNA_i^{Met}$ molecules (Feng et al. 2000).

The reverse transcription mechanism is highly error-prone due to the lack of proofreading repair activity of reverse transcriptase and RNA polymerase.

Misincorporation or insertion–deletions result in base substitutions, premature stop codons or frameshifts in the new DNA copies giving rise to divergent sequences from the canonical structure (Preston 1996). Therefore, a hallmark of retroelements is the genetic heterogeneity of individual copies belonging to a same family (Casacuberta et al. 1995). Retrotransposons are ubiquitous in the plant kingdom and structural variability within different families of plant retrotransposons has been described extensively. Several reports indicate that despite the functional importance of LTR sequences, they are one of the most rapidly evolving retrotransposon regions (Casacuberta and Grandbastien 1993; Casacuberta et al. 1995; Vicient et al. 2005). Variability within retrotransposon families also includes elements lacking coding capacity for mobility-related protein, among them solo-LTRs or tandem arrays of LTRs and internal domains resulting from intrachromosomal recombination between LTRs of the same or different inserted elements (Shirasu et al. 2000; Vicient et al. 1999). Other groups of non-autonomous LTR retrotransposons comprising two LTR flanking a non-functional central domain have also been identified in plant genomes: TRIMs (terminal-repeat retrotransposons in miniature) are short LTR elements (500 bp maximum) present in both dicotyledons and monocotyledons; *Morgane* is a 1,871 bp-long element related to the *gypsy* class from wheat that displays an internal non-functional *pol*-like domain (Sabot et al. 2006), and finally LARDs (large retrotransposons derivatives) derived from *gypsy* elements have been described in barley and rice (Kalendar et al. 2004). Although non-autonomous, TRIMs, *Morgane* and LARDs are multi-copy elements that are proposed to be *trans*-mobilized by the protein complex of intact retrotransposons (Kalendar et al. 2004). Chimerical elements resulting from recombination events between close retrotransposons families can also be observed: for example, *BARE-2* is a chimerical element related to the *BARE-1-Wis-2* retrotransposon grass family produced by events of recombination between *BARE-1* and *Wis-2a*. (Vicient 2005). Intrachromosomal recombination is not the only mechanism at the origin of non-canonical elements: template switching between the two packaged RNA molecules during retrotransposition has also been proposed as a mechanism leading to deleted copies. This mechanism has been put forward to explain the structure of *wsStonor*, an element belonging to the *Stonor* maize retrotransposon family that had suffered a 2.6 Kb-long deletion of the entire *gag* region and part of the protease domain (Marillonnet and Wessler 1998).

The grapevine genome sequence, sized 487 Mb, has recently been obtained from the highly homozygous genotype PN40024. The current annotation of the transposable elements revealed a large prevalence of retrotransposons over transposons, most of which (47%) correspond to *Ty1-copia* like retrotransposons. Distribution of transposable elements over the grape chromosomes is mainly restricted to low gene-density regions and introns (The French–Italian Public Consortium for Grapevine Genome Characterization 2007). Before the grape genome sequencing, three grapevine retrotransposons were described: a unique copy of *Vine-1*, a Ty1-*copia* like element (Verriès et al. 2000) and of *Gret 1*, a Ty3-*gypsy* like element (Kobayashi et al. 2004) and the *Tvv*1 family whose consensus sequence was reconstituted by chromosome walking (Pelsy and Merdinoglu 2002). It was shown that in a single genome, *Tvv1* elements shared a single, highly conserved 4,143 nucleotides long ORF and possessed LTRs sized between 149 and 157 bp and an untranslated leader (UTL) region, upstream of the ORF, highly variable in size and sequence (Pelsy and Merdinoglu 2002). In the present study, we have investigated the structural variability of the *Tvv1* family in the grapevine genome to determine the canonical sequence. Then, we have analyzed the structure and dispersion of two remarkable *Tvv1* elements that had suffered large deletions of their coding sequence. Based on their structural and genetic characteristics, we found strong evidence supporting our hypothesis of illegitimate recombination during reverse transcription to explain the underlying mechanism leading to the formation of these deleted elements.

Materials and methods

Plant material and DNA extraction

The two deleted *Tvv1* elements were characterized in cabernet sauvignon (certified clone 169; Anonymous 1995). Self-progenies were derived from cabernet sauvignon (Chilean clone), muscat Ottonel (certified clone 59), pinot noir (certified clone 162) and riesling (certified clone 49). Distribution of the deleted elements was evaluated in a collection of 58 *V. vinifera* accessions, represented by wine and table grapevine (52 varieties), and wild vines (6 accessions) belonging to the ampelographic collection of INRA-Colmar (France).

Young, expanded leaves were collected from individual plants and ground into fine powder with liquid nitrogen. Total genomic DNA was purified using Dneasy ™ Plant Mini-Kit (Qiagen, Hilden, Germany) as described by the supplier.

PCR conditions and fragment analysis

All primers were synthesized by MWG Biotech AG (Ebersberg, Germany); their sequences and location are given in Table 1 and Fig. 1. Long-range amplifications

Table 1 Primers used in the study

No.	Primer	Orientation	Sequence (5′–3′)	Location	Position/*Tvv1*-VB
1	P18	Sense	GTTGCTGATATAGTGCCTATTGTGTCA	ORF	939–965
2	Pltr4	Antisense	CAATTGAGAGTCCT(AG)AATTAGGAA GGT	LTR	56–82//5129–5155
3	Pltr3	Sense	CCTTCCTAATT(CT)AGGACTCTCAATTGT	LTR	57–83//5130–5156
4	Pdel1	Antisense	TT(CT)TGCACTC(AG)AACGGGATACAACACT	ORF	4701–4727
	Pvv78xfw	Sense	AAAGGTGCAGCCAACCAAACTCTCTTC	5′ Δ3460 host region	–
	GW16	Antisense	GTGGTCATCTTTAGCAACACTTCTCAA	ORF	1032–1058
	GW13	Antisense	GGACGATGAAAGGCATTCCATACATCA	3460bp-deletion	1250–1276
	Pvv78xrv	Antisense	GCAAGTAGTAAGGCAGGASTTAAGAGG	3′ Δ3460 host region	–
	Pdel4	Sense	GTAAGAAGCAAAGTGTTGTATCCCGTT	ORF	4690–4716
	GW1	Sense	CATTTGGCATGAACAAGAGCGAGAAAG	3460bp-deletion	4027–4053
	Pbo5	Sens	TGTCTATAGGCAAATGGAAGGGAACTA	5′ Δ3001 host region	–
	Pbo3	Antisense	TGGGGTGGAAACGAACGAGTGGAAATA	3′ Δ3001 host region	–
	PVB1	Sense	TGTTTCTCAT GCACATTACA AG	5′ VB host region	−132 to −110
	Pltr2	Antisense	ATCCCTAAGGAAACAACCAA	LTR	33–52//5106–5125
	PVB2	antisense	CCTTGTGAGGGATAACATAACC	3′ VB host region	+424 to +446
	Pltr1	Sense	CCTAATTCAGGACTCTCAAT	LTR	61–80/5134–5153

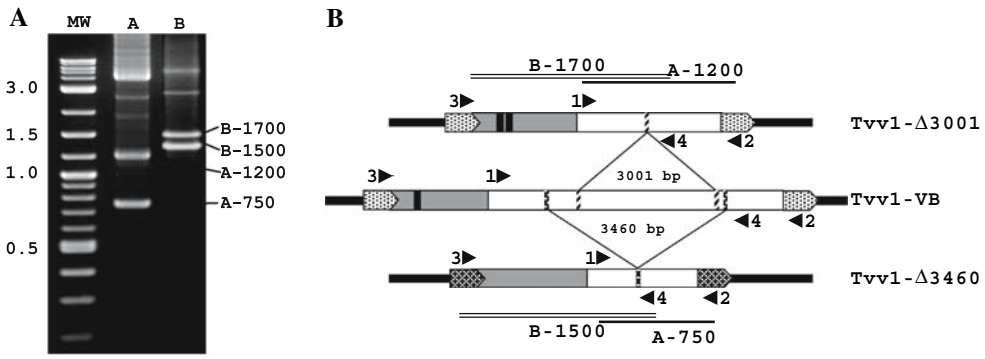

Fig. 1 Structure of the *Tvv1*-VB, *Tvv1*-Δ3001 and *Tvv1*-Δ3460 loci. **a** A and B PCR-amplifications of *T*vv1 elements of cabernet sauvignon. Molecular weight standard range from 0.2 to 10 Kb. Additional low intensity PCR products should result from amplification of additional deleted copies. **b** Schematic representations of the full-length *Tvv1*-VB of Vidal blanc and of the two deleted elements *Tvv1*-Δ3001 and *Tvv1*-Δ3460. *Arrow boxes* represent LTRs, *gray boxes* UTL and *open boxes* ORF regions. *Black arrows* represent primers used in the study, whose numbers are reported in Table 1

were performed using the primer pair Pltr4 and P18 for A amplifications and the primer pair Pltr3 and Pdel1 for B amplifications. Amplification reactions were performed in a mixture containing 20 ng DNA, 500 nM each primer, 1.5 mM $MgCl_2$, 200 μM deoxynucleoside triphosphate and 0.5 U DyNAzyme EXT™ Taq DNA polymerase (Finnzymes, Espoo, Finland) in a 20-μl reaction buffer (DyNAzyme EXT Mg-free buffer, Finnzymes). PCR cycling conditions were 94°C for 2 min, 35 cycles of 94°C for 10 s, 55°C for 30 s and 68°C for 8 min and ended by a final extension at 68°C for 7 min. *Tvv*1-VB insertion site was PCR-amplified from primers designed in the 5′ (PVB1) or 3′ (PVB2) flanking regions of *Tvv*1-VB and in the LTR, Pltr1 and Pltr2, respectively. *Tvv1*-Δ3001 flanking regions were identified by a genome walking procedure on a BAC clone consisted in two successive PCR amplifications (Pelsy and Merdinoglu 2002), the 5′ region was amplified using primers in antisense orientation GW16 and Pltr4 and the 3′ region was amplified using primers in sense orientation Pdel4 and Pltr3. *Tvv1*-Δ3460 flanking regions were amplified using primers Pvv78x designed from the pinot noir sequence upstream and downstream from the *Tvv1*-Δ3460 element, Pvv78xfw used in pair with primer GW16 and Pvv78xrv with Pdel4. Amplification reactions were performed in a mixture containing 10 ng DNA, 500 nM each primer, 150 μM deoxynucleoside triphosphate and 0.5 U GoTaq® DNA polymerase (Promega Corporation, Madison, WI, USA) in a 20-μL GoTaq® reaction buffer.

PCR cycling consisted of 7 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 followed by an elongation step of 65°C for 10 min. All amplifications were performed on a GeneAmp® PCR System 9700 thermocycler (PE Applied Biosystems, Foster City, CA, USA). Aliquot of the PCR products were analyzed by electrophoresis on a 2% agarose gel in 1 × TAE (40 mM TRIS-acetate pH 7.8, 20 mM EDTA) loading 2-Log DNA ladder as a standard of the molecular weight (New England BioLabs Inc., Ipswich, MA, USA).

For bands of interest, PCR fragments were cloned into the pCR® 4-TOPO cloning vector (InVitrogen, Groningen, NL, USA), according to the supplier instructions. Two random clones were selected for forward and reverse sequencing by PCR using the BigDye Terminator v3.1 Cycle Sequencing kit (Applied Biosystems). Samples were processed on an ABI3100 Genetic analyzer, a 16-capillary electrophoresis sequencing instrument. Computer-assisted analysis of the sequence data was performed using DNAsis 2.1 (Hitachi Software Engineering Co, Ltd) and Vector NTI 7 (InforMax Inc.). The nucleotide data of the 12 reported fragments have been deposited in GenBank database under accession numbers EU293346-EU293357. *Tvv1* elements were searched in the grapevine genome using Genoscope BLAST server (http://www.cns.fr/cgi-bin/blast_server/projet_ML/blast.pl). They were individually examined for position of their LTRs, UTL, ORFs using Vector NTI 7 (InforMax Inc.).

Results

The *Tvv1* family

The consensus sequence of *Tvv1* grapevine retrotransposons was first reconstituted by chromosome walking from the variety riesling (Pelsy and Merdinoglu 2002). A sequence homologous to an expected *Tvv1* element was identified in the BAC clone A1-16 (GenBank accession no. EU304807) from the grapevine variety *Vidal blanc* (Zyprian et al. personal communication). This clone contained a full-length *Tvv1* copy sized 5,222 bp, referenced as *Tvv1*-VB, whose structural characteristics are two perfect 149-bp long LTRs flanking an internal domain. *Tvv1*-VB LTR sequence is identical to *Tvv1* LTR L3B3 previously identified by chromosome walking (GenBank accession no. AF478367), but while the *Tvv1*-VB is flanked by perfect 5-bp direct repeats (ATTCT), the 5-bp downstream of the LTR L3B3 (TTTTC) indicate that the two insertion sites are different. The 4,924 bp-long internal domain includes the priming motifs purine binding site (PBS) and polypurine tract (PPT) described for all *Tvv1* elements and encode a putative 1,382 amino acid-long polyprotein within a single translational reading frame, uninterrupted neither by stop-codons nor by frameshifts, showing 98.9% identity to the *Tvv1* consensus sequence. UTL region, starting by the PBS complementary to the $tRNA^{met}$, is 768 bp-long, a size in the range of expected UTL found in the riesling variety (Pelsy and Merdinoglu 2002). No tri-nucleotide ATG that could indicate the starting point of a putative open reading frame was found in the UTL region. The GC content of *Tvv1*-VB is 42.5%, but it varies along the retrotransposon sequence: from 30% at the LTRs, to 41.5% at the ORF and to 53.5% at the UTL. The presence of *Tvv1*-VB at the predicted locus in the genome of *Vidal blanc* was assessed in the variety cabernet-sauvignon by a PCR-amplification using primers designed in the 5′ and 3′ flanking regions of the element and in the highly homozygous genotype PN40024 by a bioinformatic analysis. Only the empty site was amplified in cabernet sauvignon and found in the sequence of PN40024 (ML8X_scaffold_43).

Later on, 26 *Tvv1* elements flanked by 5-bp direct repeats were identified in the PN40024 sequence (Table 2). A total of 22 copies were full length; their sizes ranged from 4,640 to 7,162 bp, with LTRs from 149 to 198 bp, UTL from 179 to 1,069 bp and *gag-pol* gene from 4,139 to 6,114 bp. Four of them displayed identical LTRs and an uninterrupted single ORF, indicating that these copies were both recent and potentially functional. None of them were fully identical to *Tvv1*-VB, the only PN25 displayed a UTL region 99% similar to the *Tvv1*-VB, but differed from *Tvv1*-VB by a 8 bp deletion in the *gag-pol* region leading to a non-functional copy. Taking into consideration elements that display identical LTRs and duplication sites, it could be considered that the canonical *Tvv1* element has 149bp-long LTRs, a 4,146bp-long ORF encoding a putative 1,382 amino acid-long polyprotein within a single translational open reading frame. PN120 and PN139 have these characteristics, but differ in their UTL region sized 824 and 891 bp, and show 69.6% identity. In addition to full-length copies, 13 incomplete copies were characterized and 4 others suffering deletions, among them PN16 will be extensively described later. Twenty *Tvv1* elements were located on independent scaffolds and three scaffolds contained two elements. Identified as a canonical element, *Tvv1-VB* sequence has been considered as the reference in the following analysis.

Two deleted copies of *Tvv1* from cabernet sauvignon

To amplify *Tvv1* copies, PCR amplifications were performed using primers P18, located at the very beginning of the ORF, and Pltr4, in the consensus part of the LTR (A amplification; Table 1). Amplification of cabernet

Table 2 Sequence characteristics of *Tvv1*-VB and 26 *Tvv1* elements identified in the grapevine genome sequence

Tvv1 number	Copy state	Total length (bp)	Direct repeat 5′	LTR		UTL (bp)	gag-pol gene (bp)	Direct repeat 3′	Location in PN40024	
				(bp)	Identity (%)				Chromosome	Scaffold
VB	Full	5,222	ATTCT	149	100.0	768	4146	ATTCT	10	scaffold_43[a]
PN120	Full	5,278	GCCAT	149	100.0	824	4146	GCCAT	18	scaffold__1
PN139	Full	5,345	TTCAT	149	100.0	891	4146	TTCAT	1	scaffold_136
PN121	Full	5,525	AAGAA	150	100.0	1,069	4146	AAGAA	6	scaffold_3
PN140	Full	5,226	AGGATA	149	100.0	743	4152	AGGATA	6	scaffold_28
PN25	Full	5,215	CCTTT	149	100.0	768	4139[b]	CCTTT	8	scaffold_115
PN136	Full	5,075	GGCC**C**	151	99.3	620	4143[b]	GGCC**A**	12	scaffold_93
PN142	Full	5,256	AGAAC	149/150	99.3	803	4144[b]	AGAAC	5	scaffold_156
PN124	Full	5,268	TTATT	151	99.3	813	4143[b]	TTATT	14	scaffold_54
PN123	Full	6,387	GTTAC	198	99.3	511	5470[b]	GTTAC	7	scaffold_44
PN145	Full	7,162	**G**CCTT	149	99.3	740	6114[b]	**A**CCTT	14	scaffold_63
PN135	Full	5,351	AATGG	150	98.7	898	4143[b]	AATGG	18	scaffold_61
PN24	Full	5,532	CAATT	151	98.7	1,076	4144[b]	CAATT	14	scaffold_54
PN141	Full	5,307	TGTTC	150	98.0	843	4154[b]	TGTT**G**	7	scaffold_20
PN128	Full	5,362	A**A**ATG	169	97.6	858	4156[b]	A**T**ATG	nc	scaffold_754
PN129	Full	5,522	GAACA	168	97.0	1,032	4144[b]	GAACA	nc	scaffold_109
PN131	Full	4,958	CATT**A**	179	96.6	466	4124[b]	CATT**C**	10	scaffold_179
PN134	Full	4,640	**A**GAAG	167/168	95.2	179	4116[b]	**G**GAAG	nc	scaffold_139
PN04	Full	5,027	AACAA	172/168	93.1	511	4166[b]	AACAA	5	scaffold_2
PN137	Full	4,666	AGATG	182/194	90.7	417	3863[b]	AGATG	2	scaffold_140
PN126	Full	4,910	ATGTC	194/182	90.7	408	4116[b]	ATGTC	6	scaffold_28
PN01	Full	5,498	ACTCC	186/168	86.6	983	4152[b]	ACTCC	nc	scaffold_80
PN122	Full	5,651	ATTAT	161/189	79.4	960	4331[b]	ATTAT	16	scaffold_10
PN132	Deleted	5,116	GAAGG	151/29[c]	nc	782	4144[b]	GAAGG	7	scaffold_20
PN144	Deleted	4,259	AATTA	150/156	93.6	nc	nc	AATTC	8	scaffold_88
PN02	Deleted	2,481	ATTAC	59[c]/150	nc	nc	nc	ATTAC	18	scaffold_24
PN16	Deleted	2,057	CCCAT	161/163	95.7	1,038	685[b]	CCCAT	8	scaffold_99

Nucleotides that are different in the direct repeats flanking an element are given in bold

[a] Location of the *Tvv1*-VB empty site

[b] ORF interrupted by a stop codon

[c] LTR partially deleted

sauvignon genomic DNA creates a major product 4.3 Kb-long, in agreement with the structural organization of the full-length *Tvv*1 elements, and two additional, strongly amplified, bands sized 750 bp and 1,200 bp named A-750 and A-1200, respectively (Fig. 1). Sequencing of these additional bands showed that they could belong to two different *Tvv1* elements suffering an internal deletion. This result was confirmed by a second PCR amplification using primers Pltr3, designed in the most conserved region of the LTR, and Pdel1, located downstream of the deletions (B amplification, Fig. 1; Table 1). In addition to the expected bands around 4Kb-long, cabernet sauvignon showed two additional bands sized 1,500 and 1,700 bp, named B-1500 and B-1700, respectively. The four unexpected PCR products (bands A and B) were cloned and sequenced.

Tvv1-Δ3001

Two cloned A-1200 fragments displayed almost identical sequences (two mismatches/1,216 bp) as well as two cloned B-1700 fragments (three mismatches/1,703 bp). Sequence analysis revealed that A-1200 and B-1700 overlapped and, when compared to the sequence of *Tvv*1-VB, they displayed the same 3,001 bp-long deletion. This result led us to believe that these two bands originated from the same deleted copy that we called *Tvv1*-Δ3001. A cabernet sauvignon BAC library was screened to find a clone displaying the deleted element (Lamoureux et al. 2006). The 5′ and 3′ regions overlapping *Tvv1*-Δ3001 LTRs and host flanking regions were determined by genome walking on the BAC clone 3E11. *Tvv1*-Δ3001 is sized

2,256 bp-long and is flanked by identical 5-bp direct repeats (TGTAC). The two *Tvv1*-Δ3001 LTRs, 150 bp-long, differ by two nucleotides and show 96% identity with T*vv1*-VB LTRs (four substitutions and one single nucleotide indel; Fig. 2). Downstream of the 5′LTR, the UTL 801-bp long starts with the PBS complementary to the $tRNA^{met}$. This UTL sequence is nearly identical to the *Tvv*1-VB UTL except for the addition of 33 bp. This additional stretch, duplicated in direct orientation in the *Tvv1*-Δ3001, is already present in the T*vv1*-VB UTL. In *Tvv1*-Δ3001 UTL, the 33 bp-repeats are separated by 10 nucleotides (Fig. 2). The deletion breakpoint of *Tvv1*-3001 located in position 1,649–4,650 in *Tvv1*-VB is characterized by the presence of a 13 bp-long stretch. In the full-length *Tvv1*-VB, two nearly identical repeats in direct orientation flank the sequence corresponding to the deletion: the 3′ stretch is identical to the deletion breakpoint of *Tvv1*-Δ3001, while the 5′ stretch differs from the 3′ stretch by one mismatch (Table 3). Hypothetical translation of the remaining ORF sequences upstream of the deletion started with the regular ATG and revealed a single putative sequence of 252 amino acids in the *gag* region uninterrupted by a stop-codon (Fig. 3). The first 247 amino acids differed by 4 amino acids from the corresponding T*vv1*-VB protein, and the deletion led to the addition of 5 amino acids followed by a stop codon. Regarding the sequence downstream of the deletion, the hypothetical translation revealed a regular sequence of 134 amino acids that differs by 3 amino acids from the corresponding T*vv1*-VB region (Fig. 3).

Because of the mechanism of transposition, the two LTRs of a retrotransposon are theoretically identical at the time of their insertion (Jordan and McDonald 1998, 1999; SanMiguel et al. 1998). Thus, calculation of the base substitution rate between the two *Tvv1*-Δ3001 LTRs, 150 bp-long and differing by two mismatches, was used to evaluate the timing of insertion. We used the average base substitution rate derived from the grass *adh1-adh2* region of 6.5×10^{-9} substitution per site per year (Gaut et al. 1996), given the assumption of a uniform mutation rate, to estimate that *Tvv1*-Δ3001 inserted into its present location ≈1 million years ago.

Tvv1-Δ3460

As for *Tvv1*-Δ3001, sequence analysis of A-750 and B-1500 from cabernet sauvignon leads to the structure of

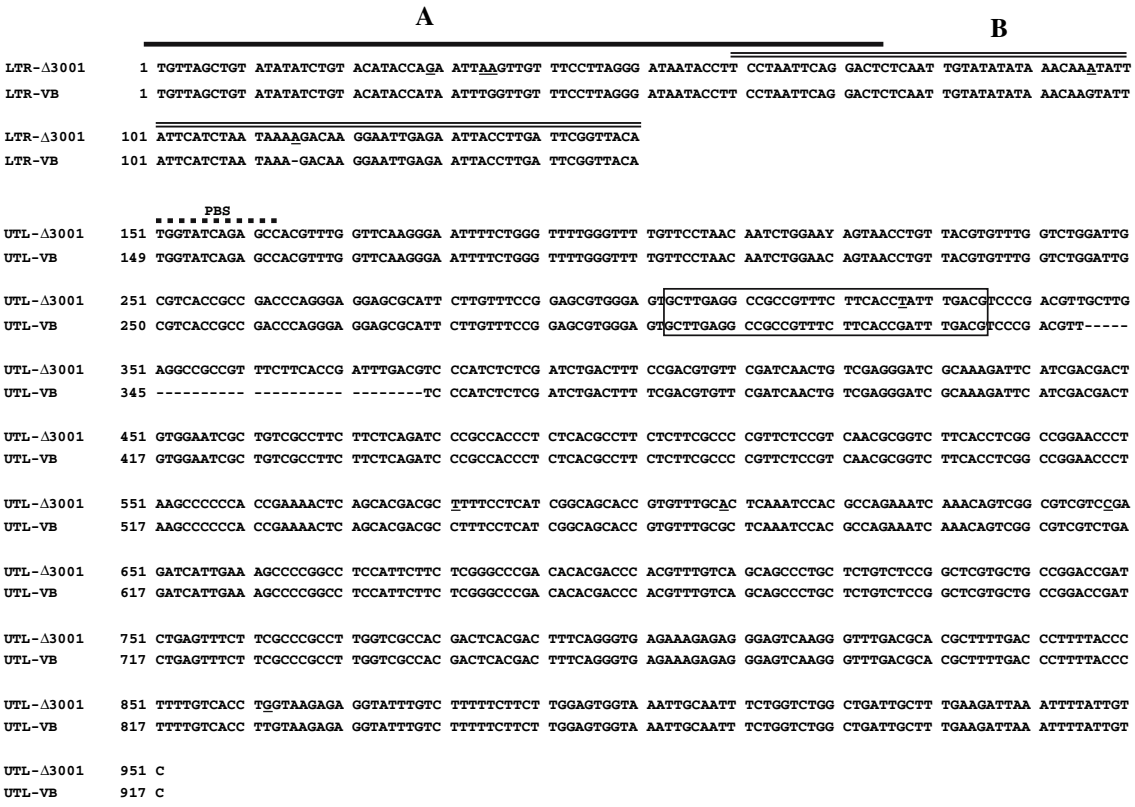

Fig. 2 LTR and UTL sequences of *Tvv1*-Δ3001 from cabernet sauvignon. Nucleotides that are different in *Tvv1*-Δ3001 and *Tvv1*-VB are *underlined*. *Solid line* (PCR A) and *double solid line* (PCR B) overlie the parts of *Tvv1*-Δ3001 revealed by PCR-amplification. The sequence duplicated in *Tvv1*-Δ3001 UTL is in *box*. *Dotted line* overlies PBS

Table 3 Size, deletion and deletion breakpoints of *Tvv1*-Δ3001 and *Tvv1*-Δ3460

	Tvv1-Δ3001 CS	*Tvv1*-Δ3460 CS
Element size (bp)	2.256	2.074
Deletion size (bp)	3.001	3,460
Deletion location/ *Tvv1*-VB	1,649–4,650	1,222–4,682
Sequence breakpoint in the deleted element	ATTGTGTGTTCTT	TGGAAAAGGAA
Homologous sequences to the	ATTGTGTGTT**T**TT	TGGAAAAGGAA
breakpoint in *Tvv1*-VB	ATTGTGTGTTCTT	TGGAAAAG**T**AA

Tvv1-Δ3460, a deleted *Tvv*1 element that had suffered a 3,460 bp-long deletion. The full sequence of *Tvv1*-Δ3460 was also found in the sequence of pinot noir (2,074 bp; accession number AM433636.2) and of PN40024 (PN16, 2057 bp; Table 2) at the same locus, giving access to the host flanking sequences of this element. Location of *Tvv1*-Δ3460 in cabernet sauvignon was confirmed by sequencing PCR products amplified from primers (Pvv78x forward and reverse) designed in the host flanking sequences and its full size deduced (2,060 bp). In the three genomes, *Tvv1*-Δ3460 element is flanked by identical 5-bp direct repeats (CCCAT) indicating a unique integration event.

In cabernet sauvignon, *Tvv1*-Δ3460 shows 5′ LTR and 3′ LTR, 165 and 161 bp-long, respectively. The two LTRs differ in the number of TA motifs in a microsatellite stretch, 11 for the 5′ LTR and 9 for the 3′ LTR, in addition to five substitutions. It is also the difference in the number of TA motifs that justify the size of *Tvv1*-Δ3460 that is different in the three genomes. The 3′LTR of *Tvv1*-Δ3460 shows 76.8% identity with the *Tvv1*-VB LTRs. Downstream of the 5′LTR, the UTL region, 1,038 bp-long, starts with the PBS complementary to the $tRNA^{met}$. *Tvv1*-Δ3460 UTL sequence shows 18.2% identity with *Tvv1*-VB UTL, differences consisting of five insertions from 138 to 5 bp-long and by many substitutions. In *Tvv1*-Δ3460, the deletion breakpoint located in position 1,222–4,682 in *Tvv1*-VB is characterized by the presence of a 11 bp-long stretch. Similar to *Tvv1*-Δ3001, two almost identical short repeats in direct orientation flank the sequence corresponding to the deletion in the full-length *Tvv1*-VB: the 5′ stretch is identical to the deletion breakpoint of *Tvv1*-Δ3001, while the 3′ stretch differs from the 5′ stretch by one mismatch (Table 3). Hypothetical translation of the ORF sequences upstream of the deletion starts with the regular ATG and reveals a single putative sequence of 113 amino acids in the *gag* region uninterrupted by stop-codon. The first 101 amino acids are nearly identical to the corresponding T*vv1*-VB protein (4 amino acids are different) and the deletion leads to the addition of 9 amino acids before being interrupted by a stop codon. Regarding the sequence downstream of the deletion, the hypothetical translation reveals a sequence of 126 amino acids that differs by 4 amino acids from the corresponding T*vv1*-VB region (Fig. 3). Cabernet sauvignon and pinot noir *Tvv1*-Δ3460 copies mainly differed in the number of TA motifs in the LTRs and also by 2 single nucleotides in the internal domain.

In the same way as for *Tvv1*-Δ3001, the timing of insertion of *Tvv1*-Δ3460 into its present location was estimated at $\geq$2.4 millions years ago, taking into consideration the five single nucleotide polymorphism but not variation in the number of repeats of the microsatellite stretch.

Segregation analysis

To determine whether *Tvv1*-Δ3001 and *Tvv1*-Δ3460 insertions are homozygous or heterozygous, transmission of B-1500 and B-1700 bands through sexual reproduction was studied. A self-progeny of cabernet sauvignon was available but consisted of only 16 individuals. To increase the number of plants, we performed the segregation analysis on 121 plants of a self-progeny of muscat Ottonel, a variety that displayed both *Tvv1*-Δ3001 and *Tvv1*-Δ3460. We previously checked that the flanking regions of *Tvv1*-Δ3001 and *Tvv1*-Δ3460 in muscat Ottonel were identical to those of other accessions, indicating that both deleted elements were at the expected locus. In both progenies, the presence versus absence of B-1700 and B-1500 was scored as segregation markers of *Tvv1*-Δ3001 and *Tvv1*-Δ3460, respectively. Eight selfed seedlings of cabernet sauvignon and 87 of muscat Ottonel displayed B-1700, while 12 selfed seedlings of cabernet sauvignon and 80 of muscat Ottonel displayed B-1500. A segregation distortion of cabernet sauvignon self-seedlings was observed, which could have probably resulted from the low number of seedlings available. The presence of *Tvv1*-Δ3001 in 78.2% of the muscat Ottonel seedlings indicated that *Tvv1*-Δ3001 segregate was a single heterozygous Mendelian locus. In the same way, the presence of *Tvv1*-Δ3460 was shown in 75% of the cabernet sauvignon and 72.1% of the muscat Ottonel self-seedlings. We assumed that only one copy of *Tvv1*-Δ3001 and one copy of *Tvv1*-Δ3460 were present in muscat Ottonel and also in cabernet sauvignon, although five positive clones containing *Tvv1*-Δ3001 were identified

Fig. 3 Comparison of the amino acid sequences of *Tvv1*-VB, *Tvv1*-Δ3001 and *Tvv1*-Δ3460. Amino acids different from *Tvv1*-VB are in *bold* and *underlined*. *Full line box* shows the nucleic acid binding domain of the *gag* protein and *dotted line boxes* motifs show I and II of the RNaseH domain. Theoretical addition of amino acids to the deleted element sequences is indicated in *italic*. For clarity, amino acids of *Tvv1*-VB absent in both deleted elements are not shown, but their number is indicated between *angle brackets*

by screening the 6X cabernet sauvignon BAC library (Lamoureux et al. 2006).

The presence of a full-length element at the *Tvv1*-Δ3001 insertion site in absence of the deleted element was investigated by amplifications using primers Pbo5 and Pbo3, designed from the host flanking regions of the cabernet sauvignon BAC clone, in 24 self-seedlings of pinot noir and of riesling, two varieties that do not display *Tvv1*-Δ3001 and in the 16 self-seedlings of cabernet sauvignon. All pinot noir and riesling self-seedlings showed a band corresponding to the expected empty site, 600 bp-long while 8 self-seedlings of cabernet sauvignon displayed a band sized 2,800 bp corresponding to the full site in agreement with the segregation of B-1700 and 12 seedlings the 600 bp band corresponding to the empty site. This result shows that the insertion site of *Tvv1*-Δ3001 is empty in absence of the deleted element. As no segregating population was available for varieties that do not displayed *Tvv1*-Δ3460, in order to reveal full-length *Tvv1* at the *Tvv1*-Δ3460 insertion site, amplifications were carried out in several varieties using primers Pvv78x designed from the host flanking regions of *Tvv1*-Δ3460 in pair with GW1 and GW13 (Table 1) designed within the 3460 bp-deletion. The results of these PCR were negative, indicating that no full-length *Tvv1* copy was present at the locus of *Tvv1*-Δ3460.

Distribution of *Tvv1*-Δ3460 and *Tvv1*-Δ3001 within a *V. vinifera* collection

Distribution of bands produced by B amplification was evaluated in a collection of 58 cultivated grapevines and wild vines. All accessions displayed B-4000 bands, most

probably indicating the presence of full-length *Tvv1* elements in their genome. Of the 58, 55 accessions amplified B-1500 and 10 accessions amplified B-1700 corresponding to *Tvv1*-Δ3460 and *Tvv1*-Δ3001, respectively (Table 4). Four accessions amplified neither B-1500 nor B-1700. Nevertheless, no accessions were found that amplified B-1700, but not B-1500.

Discussion

Canonical *Tvv1* element

Based on the structure of different full-length copies of *Tvv1* identified in the PN40024 sequence and in a BAC clone of Vidal blanc, we have deduced the structure of the canonical *Tvv1* element copy constituted by two perfect LTRs, 149 bp-long, flanking an ORF 4,146 bp-long that encode a putatively functional 1,382 amino acid-long polyprotein within a single translational reading frame. Two copies from the PN40024 genome sequence in addition to the *Tvv1*-VB were identified as canonical, the three differing in their UTL region sized 768–891 bp. The perfect identity between both LTRs indicates that these copies had recently inserted at their loci (<0.5 million years). No identical elements inserted at different loci were identified among the 22 full-length copies identified in the PN40024 genome. Thanks to the homozygocity level of this accession estimated at about 93% (The French–Italian Public Consortium for Grapevine Genome Characterization 2007), half of the *Tvv1* could be represented in this genome sequence, indicating that *Tvv1* family is constituted by a rather small number of elements, most of them being inactive.

Tvv1-Δ3001 and *Tvv1*-Δ3460, two deleted copies of *Tvv1* from cabernet sauvignon

Tvv1-Δ3001 and *Tvv1*-Δ3460 are two *Tvv*1 elements that have undergone major rearrangements compared to *Tvv1*-VB used as reference. Both elements had suffered deletions leading to the early stopping of the polyprotein synthesis before the nucleic acid binding site of the *gag* domain. *Tvv1*-Δ3001 shows a 3,001 bp-long deletion within the ORF in addition to a 33 bp-duplication within the UTL region. Another duplication, 32 bp-long, has previously been reported in the *Tvv1* fragment 3B2 at the end of the RNAseH domain (accession number AF478366; Pelsy and Merdinoglu 2002). Both duplications are in direct orientation, but while duplications of 3B2 are in tandem, ten nucleotides separate the duplication in the *Tvv1*-Δ3001 UTL. *Tvv1*-Δ3460 shows a larger deletion, 3,460 bp-long,

Table 4 Distribution of *Tvv1* full-length and deleted elements within 52 cultivated grapevine and six wild vines, *Vitis vinifera* ssp *silvestris*

Accession full name	*Tvv1* B-4000	*Tvv1-Δ3001* B-1700	*Tvv1-Δ3460* B-1500
aligoté B	+	–	+
aubin vert N	+	–	+
auxerrois B	+	–	+
bachet noir N	+	–	+
beaunoir N	+	–	+
cabernet franc N	+	–	+
cabernet sauvignon N	+	+	+
carignan N	+	–	+
chardonnay B	+	–	+
chasselas blanc B	+	–	+
chenin B	+	–	+
cinsaut N	+	+	+
clairette B	+	–	+
colombard B	+	–	+
corbeau N	+	–	+
côt N	+	–	+
folle blanche B	+	–	+
franc noir de la Haute Saone N	+	–	+
gamay blanc Gloriod B	+	–	+
gamay noir N	+	–	+
gewurztraminer B	+	–	+
gouais B	+	–	+
grenache N	+	+	+
knipperlé B	+	–	+
marsanne B	+	–	+
mauzac B	+	–	+
melon B	+	–	+
merlot N	+	–	+
mourvèdre N	+	–	–
muscat cendré B	+	+	+
muscat d'Alexandrie B	+	–	+
muscat d'Alsace rouge N	+	+	+
muscat de Hambourg N	+	–	+
muscat Ottonel B	+	+	+
muscat à petits grains B	+	+	+
muscat reine des vignes B	+	–	+
muscat de Saumur B	+	–	+
persan N	+	–	+
peurion N	+	–	+
pinot noir N	+	–	+
riesling B	+	–	+
romorantin N	+	–	+
roublot N	+	–	+
roussanne B	+	–	–
sacy B	+	–	+
sauvignon blanc B	+	+	+

Table 4 continued

Accession full name	*Tvv1* B-4000	*Tvv1-Δ3001* B-1700	*Tvv1-Δ3460* B-1500
sémillon B	+	–	–
sylvaner B	+	+	+
syrah N	+	–	+
tannat N	+	–	+
ugni blanc B	+	–	+
viogner B	+	–	+
Vitis vinifera ssp silvestris 50K	+	–	+
Vitis vinifera ssp silvestris 50L	+	–	+
Vitis vinifera ssp silvestris 53I	+	–	+
Vitis vinifera ssp silvestris 53J	+	–	+
Vitis vinifera ssp silvestris C25S2B	+	+	+
Vitis vinifera ssp silvestris C25S6	+	–	+

Cultivated grapevine types are white (*B*) or black (*N*)

encompassing the *Tvv1*-Δ3001 one. Moreover, *Tvv1*-Δ3460 UTL (1,038 bp) mainly differs from the *Tvv1*-VB UTL (768 bp) by insertions of five stretches as previously reported to explain size variation of the *Tvv1* UTL regions (Pelsy 2007).

The deleted copies of *Tvv1*-Δ3001 and *Tvv1*-Δ3460 are monogenic traits

Segregation of B-1700 and B-1500 shows that only one copy of Tvv1-Δ3001 and one of copy of Tvv1-Δ3460 are present in the muscat Ottonel and in the cabernet sauvignon genome, assuming that the observed segregation bias for B-1700 in the cabernet sauvignon self-progeny results from the low number of seedlings available. As they have not been spread around the grapevine genome, these two deleted copies cannot be considered as members of a subgroup of multicopy elements that lack protein-coding capacity such as TRIMs, *Morgane* or LARDs, whose mobility depends on trans-complementation by products of autonomous retrotransposons (Kalendar et al. 2004).

Hence, the difference observed between the sequenced A-1200, B-1700 A-750 and B-1500 fragments that were amplified from a single template must probably result from experimental errors (average mismatch 1/770 bp).

Formation of the deleted copies of *Tvv1*-Δ3001 and *Tvv1*-Δ3460 by illegitimate recombination

Retrotransposons often serve as substrates for recombination. Thanks to LTRs homology, recombination can occur between two LTRs of integrated retrotransposons through double-strand break repair. Such events lead to deletions leaving solo LTRs (Shirasu et al. 2000). In *Tvv1*-Δ3001 and *Tvv1*-Δ3460, the deletion breakpoint is marked by a stretch sized 13 and 11 bp, respectively, far shorter than common LTR sequences. The same short stretches were found almost identically in the canonical *Tvv1* flanking the deleted sequence. This feature provides an argument, albeit indirect, in favor of the deletions resulting from illegitimate homologous recombination events.

The deletion events leading to *Tvv1*-Δ3001 and *Tvv1*-Δ3460 could have occurred by recombination within one integrated *Tvv1* copy or between two independently inserted *Tvv1* copies. Research in *Arabidopsis* and tobacco (Gorbunova and Levy 1997; Kirik 2000) showed that strand rejoining after a break frequently occurs at short repeats and results in deletion from a few base pairs to several kilobases of DNA, the average deletion size depending on the species and being significantly shorter in tobacco than in *Arabidopsis* (920 vs. 1,341 bp; Kirik 2000). Such recombination events between identical stretches of two *Tvv1* copies can however be excluded because both deleted retrotransposons are flanked by perfect host repeats that mark specifically each element insertion. Recombination between identical stretches within a single *Tvv1* copy would have implied the presence of a full-length element prior to the recombination event. This second hypothesis can also be excluded because the insertion sites of *Tvv1*-Δ3001 and *Tvv1*-Δ3460 in the grapevine varieties studied either contain the deleted element or are empty.

The deletion events could also have occurred prior to integration. A source of genetic variation for viruses results in the exchange of genetic information facilitated by the presence of two RNA genomes in the same particle (Hu and Temin 1990). During retrovirus replication, the two RNA molecules packaged as a dimer can form a local heteroduplex at a region of self-complementarities, facilitating the generation of homologous recombinants. Recombination mainly occurs during original minus strand DNA synthesis and a misalignment of the two RNA molecules, thanks to repeated sequences, can result in a deleted copy (Temin 1993). In Brome mosaic virus (BMV) a nucleotide identity as short as 15 nucleotides can support efficient homologous recombination events (Nagy and Bujarski 1995). Retrotransposons such as Ty1 elements are known to recombine during retrotransposition (Boeke et al. 1986). In the same way as for virus, recombination can take place in the VLP of retrotransposons between the two packaged RNA molecules prior to the insertion of the newly synthesized DNA molecule. Such a mechanism could explain the formation of the *Tvv1* deleted copies. Supporting this idea, translation of

the sequences upstream and downstream of the deletion of *Tvv1*-Δ3001 and *Tvv1*-Δ3640 leads to the original ORF without stop codon, which is expected if the two deleted elements were formed from autonomous copies of *Tvv*1. Marillonnet and Wessler (1998) have put forward this hypothesis to explain the structure of *wxStonor*, which suffered a deletion including the full *gag* region and a portion of the protease domain of a *Stonor* element, but leaving intact the remaining ORF. However, an extensive study of all *Stonor* members showed that most of their structural variation was mainly restricted to the region just downstream of the 5′LTR. Therefore, the authors proposed that the *wxStonor* deletion occurred during retrotransposition thanks to hybridization between extra nucleotides resulting from the addition of the tRNA in the PBS region and a homologous sequence downstream of the deletion (Marillonnet and Wessler 1998). Contrary to *wxStonor*, *Tvv1*-Δ3001 and *Tvv1*-Δ3640 deletions span the ORF, leaving part of the *gag* region upstream and of the RnaseH domain downstream. In addition, the deletion breakpoints of *Tvv1*-Δ3001 and *Tvv1*-Δ3640 are characterized by 13 or 11 bp-long stretches, which are duplicated in the canonical copy at both sides of the deleted regions (Table 3). All these results lead us to propose a model explaining the generation of *Tvv1*-Δ3001 and *Tvv1*-Δ3640 that implies illegitimate recombination within one RNA molecule of a full-length *Tvv1* element or between two RNA molecules co-packaged in the VLP during reverse transcription, prior to the integration of the deleted copy into the host genome. Thanks to the two short duplicated stretches, the RNA molecule(s) could have misaligned, then led to template switching during minus strand DNA synthesis and finally to the synthesis of the deleted element. Based on the similarity of their UTL sequences, *Tvv1*-Δ3001 would derive from a copy very similar to the putative active copy *Tvv1*-VB, while *Tvv1*-Δ3460 displaying a different UTL sequence would have arisen from a different master copy of *Tvv1*.

Chronology of the *Tvv1*-Δ3001 and *Tvv1*-Δ3460 insertions

Flanking regions of *Tvv1*-Δ3460 were shown to be identical for three varieties, cabernet sauvignon, pinot noir and muscat Ottonel. This result excludes that, at least in these varieties, different *Tvv1*-Δ3460 copies were issued from independent recombination events and inserted at different sites. Hence, we assert that *Tvv1*-Δ3460 is one fixed copy resulting from a unique recombination event. Its presence in most of the varieties tested indicates that it has most likely been dispersed through natural intermixing. However, in the three varieties, *Tvv1*-Δ3460 LTRs differed in the number of TA motifs. We propose that variation in the number of microsatellite repeats occurred subsequent to integration.

The dispersion by crosses of a single-locus deleted element is supported by pedigree reconstitution: *Tvv1*-Δ3001 is present in ten varieties; some of them related as previously shown by microsatellite analysis. For example, cabernet sauvignon, a sibling from cabernet franc and sauvignon blanc (Bowers and Meredith 1997), shares *Tvv1*-Δ3001 with sauvignon blanc, and muscat cendré has inherited *Tvv1*-Δ3001 from muscat Ottonel (Bronner 2003).

Distribution of B-1500 and B-1700 is very different: while most of the varieties and wild vines (55/58) display *Tvv1*-Δ3460, only ten varieties display *Tvv1*-Δ3001. This result can be explained by the chronology of the insertions of *Tvv*1-Δ3001 ($\approx$1 million years) being more recent than *Tvv*1-Δ3460 ($\geq$2.4 millions years) that is older, more widely distributed and the sequence of which is less conserved. It is in agreement with the fact that we have not found any accessions that displayed *Tvv*1-Δ3001, but not *Tvv1*-Δ3460, and also with paleo-botanical data that record the first Vitaceae fossils from the Miocene (23.0–5.3 millions years; Galet 1988).

Acknowledgments The authors thank Dr. Eva Zyprian and Dr. Reinhardt Töpfer from the Federal Centre for Breeding Research on Cultivated Plant, Institute for Grapevine Breeding Geilweilerhof (Siebeldingen, Germany) for kindly providing the partial sequence of the BAC clone A1-16, Riccardo Velasco from the Istituto Agrario San Michele all'Adige (Italy) for the pinot sequence (VV78X150312.9) and Anne Françoise Adam-Blondon (INRA-URGV, France) for the BAC clone 3E11. We thank Lucie Bevilacqua-Schehrer and Cindy Ponama for their valuable technical assistance, Vincent Dumas and the experimental domain staff for grapevine accessions, Malek Alioua (ULP-Strasbourg, France) for sequencing, and Pere Mestre for careful revision of the English. Cedric Moisy is a doctoral fellow supported by the Conseil Régional d'Alsace. All experiments described in this study comply with the current laws of France.

References

Anonymous (1995) Catalogue des variétés et clones de vigne cultivés en France, ENTAV, Ed, Le Grau du Roi. France, p 357

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

Boeke JD, Styles C, Fink R (1986) *Saccharomyces cerevisiae SPT3* gene is required for transposition and transpositional recombination of chromosomal Ty elements. Mol Cell Biol 6:3575–3581

Bowers J, Meredith C (1997) The parentage of a classic wine grape, Cabernet Sauvignon. Nature Genet 16:84–87

Bronner A (2003) Muscats et variétés muscatées. Inventaire et synonymie universels, des origines à nos jours, Avenir-Oenologie, INRA Editions/Oenoplurimedia, p 222

Casacuberta JM, Grandbastien M-A (1993) Characterization of LTR sequences involved in the protoplast specific expression of the tobacco *Tnt1* retrotransposon. Nucl Acids Res 21:2087–2093

Casacuberta JM, Vernhettes S, Grandbastien M-A (1995) Sequence variability within the tobacco retrotransposon *Tnt1* population. EMBO J 14:2670–2678

Feng Y-X, Moore SP, Garfinkel DJ, Rein A (2000) The genomic RNA in Ty1 virus-like particles is dimeric. J Virol 74:10819–10821

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

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

Galet P (1988) Cépages et vignobles de France. Les vignes américaines, Edition, p 553

Gaut BS, Morton BR, McCaig BC, Clegg MT (1996) Substitution rate comparisons between grasses and palms: Synonymous rate differences at the nuclear gene *Adh* parallel rate differences at the plastid gene *rbcL*. Proc Natl Acad Sci USA 93:10274–10279

Gorbunova V, Levy AA (1997) Non-homologous DNA end joining in plant cells is associated with deletions and filler DNA insertions. Nucl Acids Res 25:4650–4657

Hu W, Temin HM (1990) Genetic consequences of packaging two RNA genomes in one retroviral particle: pseudodiploidy and high rate of genetic recombination. Proc Natl Acad Sci USA 87:1556–1560

Jordan IK, McDonald JF (1998) Evidence for the role of recombination in the regulatory evolution of *Saccharomyces cerevisiae* Ty elements. J Mol Evol 47:14–20

Jordan IK, McDonald JF (1999) Tempo and mode of Ty element evolution in *Saccharomyces cerevisiae*. Genetics 151:1341–1351

Kalendar R, Vicient CM, Peleg O, Amamthawat-Jonsson K, Bolshoy A, Schulman A (2004) Large retrotransposon derivatives: abundant, conserved by non autonomous retroelements of barley and related genomes. Genetics 166:1437–1450

Kirik A, Salomon S, Puchta H (2000) Species-specific double-strand break repair and genome evolution in plants. EMBO J 19:5562–5566

Kobayashi S, Goto-Yamamoto N, Hirochika H (2004) Retrotransposon-induced mutations in grape skin color. Science 304:982

Lamoureux D, Bernole A, Clainche IL, Tual S, Thareau V, Paillard S, Legeai F, Dossat C, Wincker P, Oswald M, Merdinoglu D, Vignault C, Delrot S, Caboche M, Chalhoub B, Adam-Blondon A-F (2006) Anchoring of a large set of markers onto a BAC library for the development of a draft physical map of the grapevine genome. Theor Appl Genet 113:344–356

Marillonnet S, Wessler SR (1998) Extreme structural heterogeneity among the members of a maize retrotransposon family. Genetics 150:1245–1256

Nagy PD, Bujarski JJ (1995) Efficient system of homologous RNA recombination in brome mosaic virus: sequence and structure requirements and accuracy of crossovers. J Virol 69:131–140

Nymark-McMahon MH, Beliakova-Bethell NS, Darlix J-L, Le Grice SFJ, Sandmeyer SB (2002) Ty3 integrase is required for initiation of reverse transcription. J Virol 76:2804–2816

Pelsy F (2007) Untranslated leader region polymorphism of *Tvv1*, a retrotransposon family, is a novel marker useful for analyzing genetic diversity and relatedness in the genus *Vitis*. Theor Appl Genet 116:15–27

Pelsy F, Merdinoglu D (2002) Complete sequence of *Tvv1*, a family of Ty*1 copia*-like retrotransposons of *Vitis vinifera* L., reconstituted by chromosome walking. Theor Appl Genet 105:614–621

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

Sabot F, Sourdille P, Chantret N, Bernard M (2006) *Morgane*, a new LTR retrotransposon group, and its subfamilies in wheats. Genetica 128:439–447

SanMiguel P, Gaut BS, Tikhonov A, Nakajima Y, Bennetzen JL (1998) The paleontology of intergene retrotransposons of maize. Nat Genet 20:43–45

Shirasu K, Schulman AH, Lahaye T, Schulze-Lefert P (2000) A contiguous 66-kb barley DNA sequence provides evidence for reversible genome expansion. Genome Res 10:908–915

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

Temin H (1993) Retrovirus variation and reverse transcription: abnormal strand transfers result in retrovirus genetic variation. Proc Natl Acad Sci USA 90:6900–6903

The French–Italian Public Consortium for Grapevine Genome Characterization (2007) The grapevine genome sequence suggests ancestral hexaploidization in major angiosperm phyla. Nature 449(7161):463–467

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

Vicient CM, Kalendar R, Schulman AH (2005) Variability, recombination, and mosaic evolution of the barley *BARE*-1 retrotransposon. J Mol Evol 61:275–291

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

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