Vitis **51** (2), 65–72 (2012)

A new *gypsy*-like retroelement family in *Vitis vinifera*

S. Godinho[1)], O. S. Paulo[2)], L. Morais-Cecílio[1)] and M. Rocheta[1)]

[1)] Centro de Botânica Aplicada à Agricultura, Instituto Superior de Agronomia, Universidade Técnica de Lisboa, Lisboa, Portugal

[2)] Computational Biology and Population Genomics Group, Centro de Biologia Ambiental/Departamento de Biologia Animal, Faculdade de Ciências da Universidade de Lisboa, Lisboa, Portugal

Summary

As a major part of most plant genomes, retrotransposons are distributed throughout the plant genome ubiquitously with high copy number and extensive heterogeneity. Various retrotransposon families with distinct structures differ in their distribution and roles among divergent plant species, due to unforeseen transposition activities.

We had performed *in silico* analysis of the *Vitis vinifera* 'Pinot Noir' genome to search for gypsy type retroelements homologues to the one identified in *Pinus radiata* (IFG7) and *P. pinaster* (PpRT1) and in *Quercus suber* (Corky). We intended to see the existence and structure of *gypsy*-like retroelements homologues in the *Vitis* genome as well as the existence of integration site preference. From all data and to perform a deeper analysis we chose 36 complete sequences copies in the *Vitis* genome. We used three genetic distance corrections, additional to p-distance to estimate retroelements insertion time and reverse transcriptase, integrase and LTR (Long Terminal Repeat) sequences to establish a phylogeny and to see the contributions of different regions according to the evolutionary rates. We found three elements with identical LTRs and two old elements that revealed recent and very old insertions as well as insertions inside other retroelements. Additionally, we found no preference for the integration site as shown by the different target site repeat for each element.

Key words: *Vitis vinifera*, LTR retrotransposons, *in silico*, phylogenetic analysis.

Introduction

Transposable elements (TE) are the single most abundant class of genetic material in higher eukaryotes. They are able to move (transpose) from one chromosomal location to another, therefore playing a central role in the structure, evolution and function of eukaryotic genomes (Bennetzen 2000). Three different approaches revealed that in average, 41.4 % of the grapevine genome is composed of repetitive/transposable elements (TEs) (Jaillon *et al.* 2007), a slightly higher proportion than that identified in the rice genome, which has a somewhat smaller size (Project. 2005).

Transposable elements are classified based upon their mechanism of transposition as well as by comparison of their genomic structures and sequences (Finnegan 1992). Class I elements, or retroelements, transpose via the reverse transcription of an RNA intermediate and can be divided into several subclasses commonly referred to as SINEs, LINE-like elements and long terminal repeat (LTR) retrotransposons.

LTR retrotransposons have a genomic structure that is virtually identical to that of retroviruses, and in fact they are close evolutionary relatives of retroviruses. They both contain *gag* and *pol* genes that encode a viral particle coat (GAG) and a reverse transcriptase (RT), ribonuclease H (RH), integrase (IN) to provide enzymatic activities for making cDNA from RNA and inserting it into the genome. Sometimes they have another region called chromo domain that can be responsible for the targeted integration. They differ from retroviruses because the latter encode an envelope protein that facilitates their movement from one cell to another, whereas LTR retrotransposons either lack or contain a remnant of an *env* gene and can be reinserted into the genome from which they came. In order to facilitate their transcription, transposable elements often encode their own promoter sequences. Such a "copy and paste" mechanism has been largely successful during the evolution of eukaryotes in which class I elements represent the largest portion of higher plant genomes. Because the long terminal repeats of LTR retrotransposons are synthesized from a single template during reverse transcription, they are identical at the DNA sequence level on integration. Therefore, if the nucleotide substitution rate for the host DNA polymerase is known, the relative integration time or age of the element can be estimated from the level of sequence divergence existing between an element's LTRs. The important roles of retrotransposons to modify genome size, remodel genome structure, and displace gene functions in the plant genome indicate that retrotransposons are an important driving force in genome evolution. In this work we study *in silico*, the structure, position and age of several *gypsy* retroelements with similarity to the one (*IFG*7/*PpRt*1) identified in phylogenetic distant species.

Material and Methods

Strategy to identify gypsy retrotransposons in Vitis genome: *In silico* analysis of the *Vitis vinifera* genome was performed to search *gypsy* type retroelements (Fig. 1) using reverse the retrotranscriptase sequence from the one identified in *Pinus radiata* (Kos-

Correspondence to: Dr. M. Rocheta, Centro de Botânica Aplicada à Agricultura, Instituto Superior de Agronomia, Universidade Técnica de Lisboa, Tapada da Ajuda, 1349-017 Lisboa, Portugal. Fax: +351-213653238. E-mail: rocheta@isa.utl.pt

Fig. 1: Schematic representation of *VvRet* class I retroelements (*gypsy*-like) with the coding regions *gag* and *pol* for capsid protein (CP) and protease (PR), reverse transcriptase (RT), Rnase H, integrase (INT) and chromo domain (CD) genes, respectively. Characteristic sequences like target sequence repeats (TSR), inverted repeats (IR), primer binding site (PBS), polypurine tract (PPT) and chromo domain (CD) often found in *gypsy*-like retroelements are also pointed.

sack and Kinlaw 1999), *P. pinaster* (Rocheta *et al.* 2006) and the one recently identified in *Quercus suber* named *Cork* (unpubl.).

We performed these searches in several steps. First, reverse transcriptase sequence was used as query to identify *gypsy* elements in the *Vitis* genomic sequence database. Second, the sequences of these retroelements were aligned to identify possible deletions and insertions in these elements. Third, both LTRs were identified in each element. Then, a new multiple alignment between each query sequence and all matches was established. This last step is necessary for identifying the boundaries of each element precisely and excluding fragments that cross-match elements belonging to different families. The structure of each element was finally determined on the basis of sequence homology of matched elements and structural characteristics of LTR retrotransposons, such as the presence of a primer binding site (PBS), a polypurine tract (PPT), and/or short target site duplications (TSDs) found at the site of integration. In rare cases in which two elements exhibited identical or near-identical sequence, flanking sequences were used to determine whether these were actually different elements at different genomic locations.

The main tools used in this approach were, Genoscope (http://www.genoscope.cns.fr) and GenBank (http://www.ncbi.nlm.nih.gov/) to search for retroelements, Expasy Translation (http://expasy.org/tools/dna.html) to translate all the sequences, GenBank Conserved domains (http://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi) to identify conserved regions in each protein. LTR Finder software (http://tlife.fudan.edu.cn/ltr_finder/) was used for identification of both LTR in each retroelement and the inverted and direct repeats were found manually.

Phylogenetic analysis of retrotransposons: To understand the phylogenetic relationship between the retrotransposon sequences three types of datasets were analysed: one with the complete sequences of the retroelements, a second one with only the LTRs and a third one with concatenated reverse transcriptase and integrase. The comparative analysis of datasets allows assessing if retrotransposon divergence is the result of recombination by both topology of inferred phylogenetic trees. To infer phylogenetic trees and to calculate substitutions rates of LTRs in each *gypsy* retroelement we used two types of software: ClustalX 2.0 (Thompson *et al.* 1997) for multiple alignments and TuneClustal (developed by Dr. B. G. Hall, with his permission) to find the best alignment with a final manual adjustment. Phylogenetic trees were inferred with maximum likelihood method (ML) (Felsenstein 1988) implemented with PAUP software (Swofford 2000) and with a heuristic search with 100 random sequence additions and a tree-bisection-reconnection. Modeltest 3.7 software (Posada and Crandall 1998) associated with PAUP was used to select the most appropriate evolutionary model for the three data sets, according to the Akaike information criterion. The data was resampled 1000 times using the non-parametric bootstrap technique to evaluate the robustness of the nodes of the phylogenetic trees. A Bayesian (BI) inference was carried out using MrBayes v3.1 (Ronquist and Huelsenbeck 2003) After some trial runs, the conditions for the Bayesian analysis were set up to ensure that the likelihood scores of the trees reached stationarity over the course of the sampling. For each analysis, a total of $1.5x10^6$ generations were implemented, with successive samples separated by 100 generations after an initial "burn in" period of 5 % of the number of samples. The Bayesian posterior probabilities (BPP) were estimated by a Metropolis-Coupled, Markov Chain Monte Carlo sampling algorithm (MCMCMC). For each dataset the model selection was carried out with MrModeltest v2.2 (Nylander 2004) and implemented according to the authors recommendations.

Estimation of the time of retrotransposon insertion: For each complete retrotransposon we have searched for copy number, localization, features and orientation inside the *Vitis* genome (www.genoscope.cns.fr).

In order to date insertion events of the copies from our database, we analyzed the LTR nucleotide divergence rate of the retrotransposons. This method was first used to date the insertion events of LTR retrotransposons in maize (SanMiguel *et al.* 1998) and subsequently extended to other species (Jordan and McDonald 1998, Bowen and McDonald 2001, Jiang *et al.* 2002) and to human endogenous retroviruses (HERVs). All the elements analyzed are intact, considering the presence of two LTRs, integrase and reverse transcriptase.

To estimate the time of retrotransposons insertion we had used four corrections for sequence divergence. The Kimura 2-parameter (K2P) distance (Kimura 1980) method allows to infer evolutionary distances based on a model of evolution in which transitions and transversions may occur at different rates. Typically the rate of transitional nucleotide substitution is higher than that of transversional substitution. With the NC model, the expected number of nucleotide substitutions between two sequences is directly estimated (*i.e.* without any correction) from the observed proportion of different nucleotides between two sequences (the so-called p-distance). The HKY model (Hasegawa *et al.* 1985) incorporates multiple parameters to create a more realistic simulation of how nucleotide sequences essentially behave. It also assumes two different rates as the K2P but contrary to the latter it also assumes unequal base frequencies (*i.e.* each base could have a frequency different form 25 %). The TVM-G model has been established by PAUP when building the LTR phylogenetic tree. The first two methods were performed in MEGA 4.1 software

(Tamura *et al.* 2007). The last methods were performed in PAUP v4.0.b4a (Swofford 2000). To calculate each retroelement insertion time the formula T=D/2*S was used where T is the retroelement age since the moment of its insertion, D is divergence between LTRs calculated by the different methods mentioned above and S is the substitution rate (Vitte *et al.* 2007). It was considered a substitution rate of $1{,}3x10^{-8}$ common to Angiosperms (Vitte and Bennetzen 2006).

Results and Discussion

IGF7 *Vitis vinifera* homologs: Genomic complexity is not just a matter of the number of different sequences, but also of the variability in their arrangement and stability. To understand how retroelements contribute to *Vitis* genome organization and evolution a search of *gypsy* type retroelements was performed, homologues to the one identified in *Pinus radiate* (*IFG*7), (Kossack and Kinlaw 1999), *Pinus pinaster* (*PpRT*1) (Rocheta *et al.* 2006) and *Quercus suber* (*Corky*) (unpubl.). From 120 retrotransposon sequences identified with homology to *IFG*7 (isolated from *P. radiata*) reverse transcriptase, 36 complete copies were chosen to be analyzed (Tab. 1).

They have a random distribution in all *V. vinifera* chromosomes and all range from 5000 to 6000 bp, except *VvRet*22 with 4960 bp and *VvRet*11 with 7193 bp. The majority of retroelements was found to have a region named chromo domain just before 3'LTR. Only three out of thirty-six has no chromo domain (Tab. 1). This domain is a conserved region with 50 amino acids found in a variety of chromosomal proteins, which appear to play a role in the functional organization of the eukaryotic nucleus and direct integration of retrotransposons to heterochromatin (Eissenberg 2001, Gao *et al.* 2008). Experimental evidence implicates the chromo domain in the binding activity of these proteins to methylate histone tails and possibly RNA (Nielsen *et al.* 2002). The four retrotransposons that insert into genes have the chromo domain, however, these insertions are in non coding regions (Fig. 5). Perhaps selective targeting to heterochromatin provided by chromo domains can be favorable for mobile elements by allowing them to avoid negative selection arising from their insertion into coding regions (Gao *et al.* 2008). The role of the chromo domain in retrotransposon insertion has not been tested experimentally, so the targeting activity remains a speculation. .

Two out of 36 didn't have Target Size Duplications (TSD) (Tab. 1). LTR size varies from 289 to 584 bp that is dependent on insertion time, as well as its identity. Twenty one out of 36 have identical LTRs in size. Additionally, there is no preference for the integration site as shown by the different target site repeat for each element although some target sites are repeated three times, "ccaac" (Tab. 1).

Phylogenetic analyses: The comparative analysis between the complete retroelements sequences, only LTRs sequences and concatenated reverse transcriptase and integrase, allows us to infer about each retroelements region contributions in a phylogenetic analysis.

Several alignments were performed with all sequences with variable win increasing gap penalties, each alignment was scored with TuneCulstal and the highest scored obtained for each dataset was chosen as the best alignment. However for each LTR pair the alignments were relatively straightforward due to little variation. As a measure of the relative goodness of fit of a statistical model we use the Akaike Information Criterion, selected for the complete retrotransposon sequences dataset the GTR+I+G model

Table 1

Retrotransposon characterization in the *Vitis* genome

Name	Size (bp)	Chromo domain	LTR 5'	LTR 3'	Target Site Repeat	Name	Size (bp)	Chromo domain	LTR 5'	LTR 3'	Target Site Repeat
VvRet 1	5167	+	360	360	attct	*VvRet* 19	5726	+	505	505	ccaac
VvRet 2	5167	+	360	360	tttaa	*VvRet* 20	5775	+	530	530	aagag
VvRet 3	5168	+	360	360	gagaa	*VvRet* 21	5776	+	529	529	cttgt
VvRet 4	5163	+	360	357	gtt	*VvRet* 22	4960	+	462	505	cccaa
VvRet 5	5166	+	360	360	acaat	*VvRet* 23	5773	+	527	527	-
VvRet 6	5168	+	360	360	acacc	*VvRet* 24	5602	+	524	524	gggag
VvRet 7	5197	+	391	391	tccc	*VvRet* 25	5773	+	529	529	tgaga
VvRet 8	5206	+	392	392	ccatg	*VvRet* 26	5684	+	463	415	actattt
VvRet 9	5193	+	392	388	gagta	*VvRet* 27	5851	-	561	580	aaacc
VvRet 10	5195	+	392	392	tatat	*VvRet* 28	5777	+	530	529	atagc
VvRet 11	7193	-	435	435	caggt	*VvRet* 29	5749	+	505	529	aaggt
VvRet 12	5221	+	422	422	aaag	*VvRet* 30	5677	+	481	481	ttccc
VvRet 13	5081	+	290	289	taatg	*VvRet* 31	5610	+	370	370	-
VvRet 14	5078	+	352	343	caac	*VvRet* 32	5698	+	506	481	agttc
VvRet 15	5024	+	352	343	agcct	*VvRet* 33	5885	+	584	584	attcc
VvRet 16	5776	+	530	529	gtcta	*VvRet* 34	5792	+	526	550	ggcct
VvRet 17	5614	+	481	481	gcaac	*VvRet* 35	5677	+	528	550	caata
VvRet 18	5776	+	529	529	tttaa	*VvRet* 36	5877	-	582	581	ggaat

(A-C transversion rate of 1.6287, G-A transition rates of 3.9157, A-T rate of 0.9329, C-G rate of 0.8345 and C-T rate of 3.6684, the proportion of invariable sites of zero and the shape parameter of the gamma distribution was 2.4261). For the LTR sequences the TVM+G model (A-C transversion rate of 1.3043, G-A transition rates of 4.3026, A-T rate of 0.5179, C-G rate of 1.0446 and C-T rate of 4.3026, the proportion of invariable sites of zero and the shape parameter of the gamma distribution was 2.9415). Finally, for the RVT and Integrase concatenated sequences the TVM+I+G model (A-C transversion rate of 1.4504, G-A transition rates of 3.2659, A-T rate of 1.0583, C-G rate of 0.9586 and C-T rate of 3.2659, the proportion of invariable sites of zero and the shape parameter of the gamma distribution was 1.7000).

We have used the Maximum likelihood tree for the complete retrotransposon dataset (Fig. 2). The two main clades (A and B) show a considerable differentiation and are well supported by both the bootstrap and the Bayesian

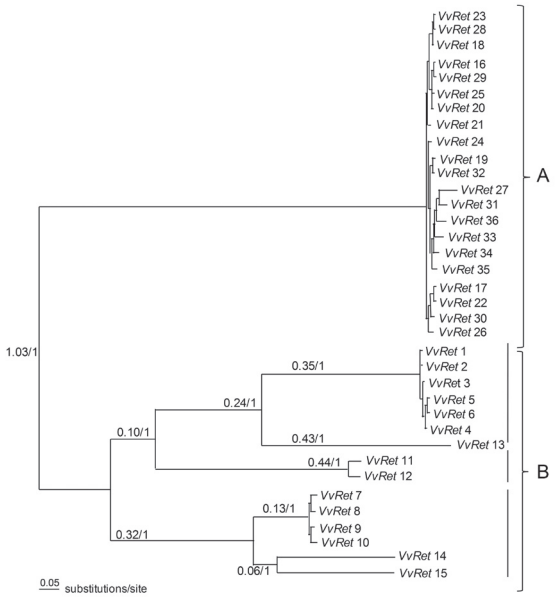

Fig. 2: Maximum likelihood phylogenetic tree obtained from the complete sequence of the retroelements rooted by midpoint. Numbers close to the branches are the bootstrap support values obtained from 1000 pseudoreplicates from maximum likelihood and the Bayesian posterior probability.

posterior probability (BPP) values. The clade B can be subdivided in three well supported sub-clades.

However, sequence analyses of the more rapidly evolving LTRs are better suited for the characterization of phylogenetic substructure within families of LTR retrotransposons. When all the LTR sequences were considered, the maximum likelihood phylogenetic tree obtained shows the same topology as the previous one, with again two main clades (Fig. 3). However, the main clade A and B can be subdivided in two and three well supported subclades, respectively. Despite the fact that the tree topology is the same, the arrangement of retroelements has some differences. This can be due to the mechanism of replication that increases the differences between both LTRs. Probably the detected clades and subclades that reflect sequence differences correspond to different insertion episodes that

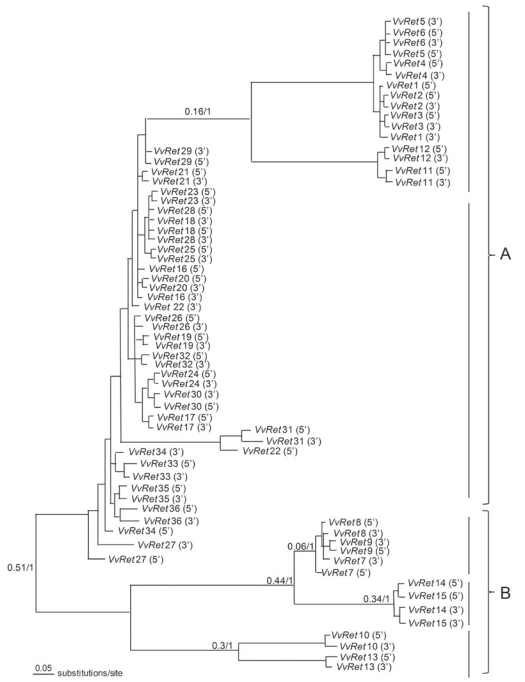

Fig. 3: Maximum likelihood phylogenetic tree derived from the LTR sequences and rooted by midpoint. Numbers in brackets indicate the LTR side. Numbers close to the branches are the bootstrap support values obtained from 1000 pseudoreplicates from maximum likelihood and the Bayesian posterior probability.

had occurred over the time. To confirm if these two datasets have a different behaviour from regions that slowly evolve, such as the reverse transcriptase (RT) or integrase (Int) (XIONG and EICKBUSH 1990) we have also performed a phylogenetic analyse with RT and Int concatenated with Mr. Bayes program and with each partition analysed with not linked parameters (Fig. 4). The concatenation of two or more regions increases the statistical power of the analysis and diverse (but not all) single-gene discrepancies are

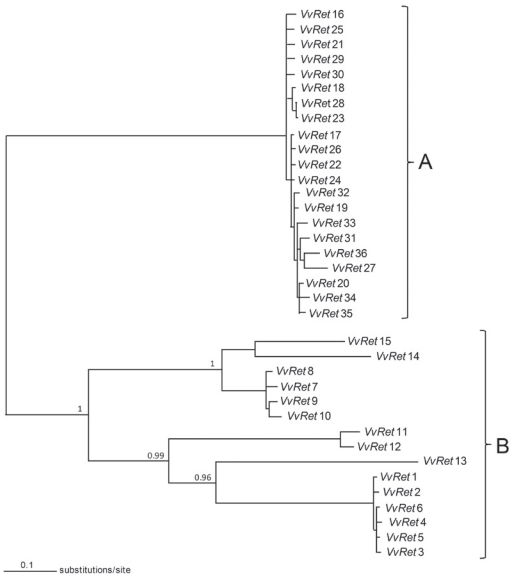

Fig. 4: Bayesian phylogenetic tree derived from the reverse transcriptase and integrase sequences concatenated. The best tree out of 150000 trees is given in MrBayes software rooted by midpoint. Numbers close to the branches are the Bayesian posterior probability.

corrected. This strategy provides an accurate perspective on phylogenetic patterns of each family. Unexpectedly, we found the same tree topology already present in Fig. 2. The more rapid evolution of LTRs sequences present in complete retroelements analysis does not have influence in tree topology, comparatively with a slowly evolving region.

Location and structure of *V. vinifera* retroelements: Given the abundance of retrotransposition events, the question arose as to how they might influence the expression of genes. Indeed, retrotransposons are involved in generating mutations through insertions near or within genes and affect their expression, usually in a negative fashion by decreasing or abolishing transcription of a gene or by detrimental alterations in transcript processing and/or stability. Retrotransposons inserted in or near plant genes have been reported in maize, rice, lettuce, wheat, tomato, tobacco, potato, and bell pepper (Kumar and Bennetzen 1999).

Here, we search for gene homologies nearby the 36 retroelements and the results reveal 15/36 (41.6 %) of LTR retrotransposon sequences lie within or near grape genes over the region of the genome analyzed in this study (Tab. 2, Fig. 5).

Four out of thirty-six had disrupted known genes. Additionally, *VvRet*17 and *VvRet*25 have a nested structure (retrotransposon inside retrotransposon) and *VvRet*36 has a copy of itself in the same strand but in an opposite direction (Fig. 6). Insertion of an LTR retrotransposon into an LTR retrotransposon would usually eliminate the target element as a potential competitor for future amplification.

In light of the recent advancements in the understanding of genome size and structure evolution, it is now presumed that genome size is a function of both genome expansion and contraction forces (Bennetzen *et al.* 2005; Devos *et al.* 2002). We found two retrotransposons, *VvRet*23 and *VvRet*31 (Tab. 1) without TSD. This fact could be an evidence of a retrotransposon removal mechanism through intra-strand recombination between LTRs of retroelements

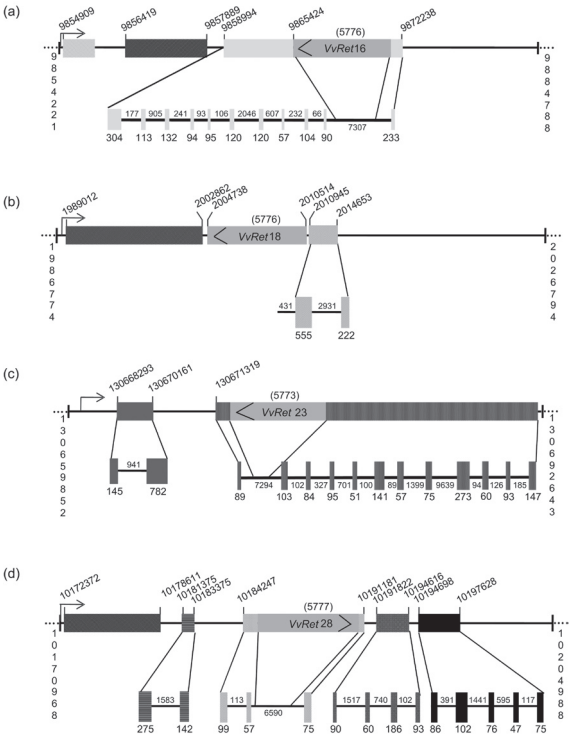

Fig. 5: Schematic representation of four retrotransposon insertion in *Vitis* genes (**a**) *VvRet1*6 , chromosome 17, scaffold 12; (**b**) *VvRet*18, chromosome 1, scaffold 84; (**c**) *VvRet*23, chromosome unknown, scaffold 77; (**d**) *VvRet*28, chromosome 11, scaffold 14. The arrow indicates the direction of transcription. Arrow head inside retroelement indicate orientation of transcription. Numbers at the beginning and the end of each scheme indicates the position in the chromosome. Numbers are in base pairs. Exons are in boxes and introns in lines. Numbers above retrotransposons indicate their size. - Incomplete retrotransposon; - protein of unknow function; - ROK protein family; - fatty acid hydroxylase superfamily; - PB1_UP2, uncharacterized protein; - Serine/Threonine protein kinases; - biotin carboxylase; - phosphate synthase; - ligase; - biotin carboxylase.

Table 2

Retrotransposon insertion in the *Vitis* genome

Name	Insertion	Gene position*	Homology	GI
VvRet 1	*CXE Carboxylesterase* gene	≈ 130 bp dws	*A. deliciosa*	82697971
VvRet 2	hAT family dimerisation domain	≈ 1540 bp dws	*O. sativa*	77553992
VvRet 4	*nbs-lrr* resistance gene	≈ 660 bp ups	*P. trichocarpa*	224096788
VvRet 8	Ribosomal protein S15 family protein	≈ 560 bp ups	*A. thaliana*	30699526
VvRet 15	nucleic acid binding / zinc ion binding	≈ 2880 bp dws	*A.thaliana*	145323089
VvRet 16	*Pantothenate kinase* gene	Inside gene	*A. thaliana*	79326098
VvRet 18	Sterol desaturase family gene	Inside gene	*S. demissum*	53793724
VvRet 22	ABC transporter	≈ 3695 bp dws	*P. trichocarpa*	224113069
VvRet 23	*GSK*-3-like gene	Inside gene	*M. sativa*	24637171
VvRet 24	FF domain-containing protein (splicing factor)	≈ 1590 bp dws	*A. thaliana*	30685515
VvRet 26	Retrotransposon	≈ 706 bp dws	*O. sativa*	77552433
VvRet 28	*Biotin carboxylase* gene	Inside gene	*R. communis*	223541050
VvRet 29	Senescence-related protein	≈ 2740 bp dws	*C. sinensis*	198400319
VvRet 33	DNA binding	≈ 1730 bp dws	*A. thaliana*	15230199
VvRet 34	*Ycf*2 gene	≈ 2880 bp dws	*C. papaya*	167391849

* bp dws - base pairs downstream retrotransposon sequence; bp ups – base pairs upstream retrotransposon sequence.

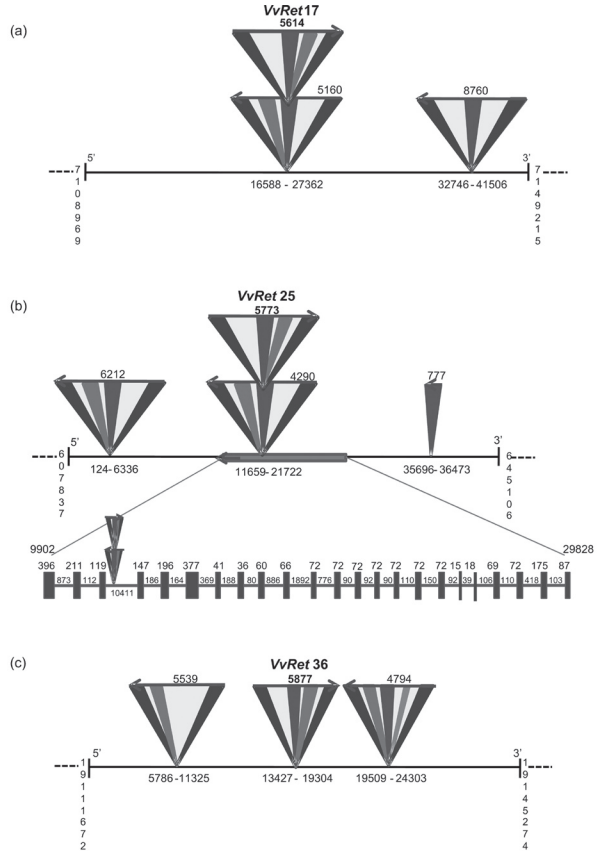

Fig. 6: Schematic representation of some *IGF*7 homologous retroelements in the *Vitis* genome (**a**) *VvRet*17 insert in chromosome unknown, (**b**) *VvRet*25 insert in chromosome 19, (**c**) *VvRet* 36 chromosome 4. Retrotransposons whose sequences were analyzed in detail in this work are marked in bold. The arrow indicates the direction of transcription. Numbers in the beginning and the end of each scheme indicates the position on the chromosome. - gene; - LTR; - zinc knuckle; - RVT; - integrase; - exon; - introns.

(Devos *et al.* 2002, Vitte and Panaud 2003). It has been shown that retroelements insertions into the coding or promoter regions of genes modulate gene function by regulation of gene expression or by formation of non-functional proteins (Hori *et al.* 2007, Xiao *et al.* 2008). Insertions of retrotransposons into the introns could be less deleterious yet not inconsequential. We found four intronic insertions (Fig. 5) that could cause splicing alterations and differential transcript accumulations (Tighe *et al.* 2002).

Our search to find insertions near or inside genes revealed fifteen retroelements close and inside genes (Tab. 2) despite they carry putative chromodomains. The presence of this protein domain is associated with the capability of these retroelements insert into heterochromatic regions (Gao *et al.* 2008). The massive presence of these elements near or inside genes could indicate an active regulation of these genes through epigenetic mechanisms associated with chromatin configuration. While these numbers are likely to change somewhat as the grape genome is better annotated, these preliminary estimates moreover indicate the potential contribution of LTR retrotransposon sequences to the evolution of gene structure and function in *Vitis* may be significant, as has also been demonstrated for the *Gret*1 insertion in a *Myb* gene controlling colour grape (Kobayashi *et al.* 2004)

Aging of the LTR-retrotransposons: Of the 36 full-length *IGF*7-like elements that we have identified in *V. vinifera*, 14 (39 %) have > 99 % LTR similarity with 3 of these (8.3 %) being completely equal (data not shown). Identical or almost identical LTRs imply that the elements have inserted recently and have not had time to accumulate mutations between LTRs. The remaining 22 (61 %) have relatively low levels of nucleotide divergence (< 99 %). Among them, 6 (16.7 %), 3 (8.3 %), 5 (13.8 %), and 8 (22.2 %) fell into different ranges of LTR similarities of 98-99 %, 97-98 %, 95-97 %, and < 95 %, respectively (data not shown).

In order to convert LTR nucleotide divergence into dates of insertion events, a substitution rate is needed for each retroelement. However, as copies have inserted at different time and different genomic locations, a global rate is difficult to estimate, and such data were not available for these retrotransposon families. To estimate the insertion times we used the average substitution rate common to Angiosperms ($1.3x10^{-8}$ substitutions per synonymous site per year), (Vitte and Bennetzen 2006).

The insertion times obtained go from 0 to 5.42, 4.92, 4.40 and 4.36 million years, with K2P, p-distance, HKY and TVM+G statistical methods, respectively (Tab. 3).

The LTRs of two retrotransposons, *VvRet* 22 and *VvRet* 27 are five and two and a half million years old, respectively. They displayed atypically high levels of sequence divergence indicating that these elements are exceptionally old or possibly that these elements are, in fact, hybrid elements generated by homologous recombination or some other recombination process. Indeed, such inter-element recombination events have been previously documented in yeast (Jordan and McDonald 1998; Jordan and McDonald 1999). Thus, we have no direct evidence that any of the full-length grapevine LTR retrotransposons analyzed in this study were generated by recombination. Moreover, two elements, *VvRet*16 and *VvRet*28 have LTRs with one nucleotide insertion and in all statistical treatments (Tab. 3) appear with no age that means a recent insertion. Grapevine is a domesticated species that has an asexual process of multiplication. Over the last 50 years it has undergone drastic reduction of diversity, owing to the restricted use of only few cultivars for the globalized wine companies. Grapevine transposable elements have no meiosis possibility to be removed from the genome and will be accumulated. These particular elements still active in the *Vitis* genome could contribute to new genotypes in the next generations.

Acknowledgements

We thank “Fundação para a Ciência e Tecnologia”, for the post-doc grant SFRH/BPD/64905/2009 to M.R. This research received no specific grant from any funding agency, commercial or not-for-profit sectors.

Table 3

Statistical models used to determine retrotransposon aging. Kimura 2-parametres (K2P); No correction (p-distance); Hasegawa, Kishino and Yano (HKY) and TVM+G. It was considered a substitution rate of $1.3x10^{-8}$ common to Angiosperms

	Satistical models							
	K2P K	TKp (Mya)	p-distance K	$1.3x10^{-8}$ TKp (Mya)	HKY K	TKp (Mya)	TVM+G K	TKp (Mya)
VvRet 1	0	0.00	0.008	0.31	0.008	0.31	0.00837	0.32
VvRet 2	0	0.00	0	0.00	0	0.00	0.00000	0.32
VvRet 3	0.008	0.31	0.008	0.31	0.008	0.32	0.00836	0.00
VvRet 4	0.008	0.31	0.008	0.31	0.008	0.33	0.00843	0.32
VvRet 5	0.006	0.23	0.006	0.23	0.005	0.21	0.00558	0.32
VvRet 6	0	0.00	0	0.00	0	0.00	0.00000	0.21
VvRet 7	0.026	1.00	0.026	1.00	0.026	1.00	0.02593	0.00
VvRet 8	0.018	0.69	0.018	0.69	0.018	0.7	0.01800	1.00
VvRet 9	0.016	0.62	0.015	0.58	0.015	0.6	0.01554	0.69
VvRet 10	0.034	1.31	0.033	1.27	0.034	1.32	0.03297	0.60
VvRet 11	0.038	1.46	0.037	1.42	0.038	1.47	0.03849	1.27
VvRet 12	0.019	0.73	0.019	0.73	0.019	0.74	0.01917	1.48
VvRet 13	0.042	1.62	0.04	1.54	0.041	1.61	0.03929	0.74
VvRet 14	0.012	0.46	0.012	0.46	0.011	0.45	0.01148	1.51
VvRet 15	0.012	0.46	0.012	0.46	0.011	0.45	0.01148	0.44
VvRet 16	0	0.00	0	0.00	0	0.00	0.00000	0.00
VvRet 17	0.006	0.23	0.006	0.23	0.006	0.24	0.00620	0.24
VvRet 18	0	0.00	0	0.00	0	0	0.00000	0.00
VvRet 19	0.006	0.23	0.006	0.23	0.005	0.23	0.00590	0.23
VvRet 20	0.006	0.23	0.006	0.23	0.005	0.22	0.00564	0.22
VvRet 21	0.006	0.23	0.006	0.23	0.005	0.22	0.00565	0.22
VvRet 22	0.141	5.42	0.128	4.92	0.114	4.40	0.11894	4.57
VvRet 23	0.004	0.15	0.004	0.15	0.004	0.16	0.00420	0.16
VvRet 24	0.006	0.23	0.004	0.15	0.005	0.22	0.00566	0.22
VvRet 25	0.004	0.15	0.004	0.15	0.003	0.15	0.00376	0.14
VvRet 26	0.015	0.58	0.014	0.54	0.009	0.37	0.00950	0.37
VvRet 27	0.07	2.69	0.066	2.54	0.068	2.64	0.07060	2.72
VvRet 28	0	0.00	0	0.00	0	0.00	0.00000	0.00
VvRet 29	0.01	0.38	0.01	0.38	0.01	0.39	0.00994	0.38
VvRet 30	0.017	0.65	0.017	0.65	0.01	0.65	0.01659	0.64
VvRet 31	0.028	1.08	0.027	1.04	0.027	1.07	0.02830	1.09
VvRet 32	0.008	0.31	0.008	0.31	0.008	0.32	0.00828	0.32
VvRet 33	0.023	0.88	0.022	0.85	0.022	0.87	0.02291	0.88
VvRet 34	0.019	0.73	0.019	0.73	0.019	0.74	0.01936	0.74
VvRet 35	0.015	0.58	0.015	0.58	0.015	0.60	0.01542	0.59
VvRet 36	0.045	1.73	0.043	1.65	0.044	1.72	0.04526	1.74

Declaration of Interests

The authors declare that they have no conflict of interests.

References

Bennetzen, J. F.; 2000: Transposable element contributions to plant gene and genome evolution. Plant Mol. Biol. **42,** 251-269.

Bennetzen, J. L.; Ma, J.; Devos, K. M.; 2005: Mechanisms of recent genome size variation in flowering plants. Ann. Bot. **95**, 127-132.

Bowen, N. J.; McDonald, J. F.; 2001: *Drosophila* euchromatic LTR retrotransposons are much younger than the host species in which they reside. Genome Res. **11**, 1527-1540.

Devos, K. M.; Brown, J. K. M.; Bennetzen, J. L.; 2002: Genome size reduction through illegitimate recombination counteracts genome expansion in *Arabidopsis*. Genome Res. **12**, 1075-1079.

Eissenberg, J. C.; 2001: Molecular biology of the chromo domain: an ancient chromatin module comes of age. Gene **275**, 19-29.

Felsenstein, J.; 1988: Phylogenies from molecular sequences: inference and reliability. Annu. Rev. Genet. **22**, 521-565.

Finnegan, D. J.; 1992: Transposable elements. Curr. Opin. Genet. Dev. **2**, 861-867.

Gao, X.; Hou, Y.; Ebina, H.; Levin, H. L.Voytas, D. F.; 2008: Chromodomains direct integration of retrotransposons to heterochromatin. Genome Res. **18**, 359-369.

Hasegawa, M.; Kishino, H.Yano, T.; 1985: Dating the human-ape splitting by a molecular clock of mitochondrial DNA. J. Mol. Evol. **22**, 160-174.

Hori, Y.; Fujimoto, R.; Sato, Y.Nishi, T.; 2007: A novel *wx* mutation caused by insertion of a retrotransposon like sequence in a glutinous cultivar of rice (*Oryza sativa*). Theor. Appl. Genet. **115**, 217-224.

Jaillon, O.; Aury, J. O.; Noel, B.; Policriti, A.Clepe, C.; 2007: The grapevine genome sequence suggests ancestral hexaploidization in major angiosperm phyla. Nature **449**, 463-467.

Jiang, N.; Bao, Z.; Temnykh, S.; Cheng, Z.; Jiang, J.; Cheng, Z.; Jiang, J.; Wing, R. A.; McCouch, S. R.; Wessler, S. R.; 2002: Dasheng: A recently amplified nonautonomous long terminal repeat element that is a major component of pericentromeric regions in rice. Genetics **161**, 1293-1305.

Jordan, I. K.; McDonald, J. F.; 1998: Evidence for the role of recombination in the regulatory evolution of *Saccharomyces cerevisiae Ty* elements. J. Mol. Evol. **47**, 14-20.

Jordan, I. K.; McDonald, J. F.; 1999: Phylogenetic perspective reveals abundant *Ty*1/*Ty*2 hybrid elements in the *Saccharomyces cerevisiae* genome. Mol. Biol. Evol. **16**, 419-422.

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

Kobayashi, S.; Goto-Yamamoto, N.; Hirochika, H.; 2004: Retrotransposon-Induced Mutations in Grape Skin Color. Science **304**, 982.

Kossack, D. S.; Kinlaw, C. S.; 1999: IFG, a *gypsy*-like retrotransposon in Pinus (*Pinaceae*), has an extensive history in pines. Plant Mol. Biol. **39**, 417-426.

Kumar, A.; Bennetzen, J. L.; 1999: Plant Retrotransposons. Annu. Rev. Genet. **33**, 479-532.

Nielsen, P. R.; Nietlispach, D.; Mott, H. R.; Callaghan, J.; Bannister, A.; Kouzarides, T.; Murzin, A. G.; Murzina, N. V.; Laue, E. D.; 2002: Structure of the HP1 chromodomain bound to histone H3 methylated at lysine 9. Nature **416**, 103-107.

Nylander, J. A. A.; 2004: MRMODELTEST version 2.2. Program Distributed by the Author. Evolutionary Biology Centre, Uppsala University.

Posada, D.; Crandall, K. A.; 1998: Modeltest: testing the model of DNA substitution. Bioinformatics **14**, 817-818.

Project., I. R. G. S.; 2005: The map-based sequence of the rice genome. Nature **436**, 793-800.

Rocheta, M.; Cordeiro, J.; Oliveira, M.; Miguel, C.; 2006: *PpRT*1: the first complete *gypsy*-like retrotransposon isolated in *Pinus pinaster*. Planta **225**, 551-562.

Ronquist, F.; Huelsenbeck, J. P.; 2003: Mrbayes 3: Bayesian phylogenetic inference under mixed models. Bioinformatics **19**, 1572-1574.

SanMiguel, P.; Gaut, B. S.; Tikhonov, A.; Nakajima, Y.; Bennetzen, J. L.; 1998: The paleontology of intergene retrotransposons of maize. Nat. Genet. **20,** 43-45.

Swofford, D. L.; 2000: PAUP*. Phylogenetic Analysis Using Parsimony (*and Other Methods). Version 4.0. Sinauer Associates, Sunderland, Massachusetts.

Tamura, K.; Dudley, J.; Nei, M.Kumar, S.; 2007: MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. Mol. Biol. Evol. **24**, 1596-1599.

Tighe, P. J.; Stevens, S. E.; Dempsey, S.; Deist, F. L.; Rieux-Laucat, F. *et al.*; 2002: Inactivation of the *FAS* gene by *Alu* insertion: retrotransposon in an intron causing splicing variation and lymphoproliferative syndrome. Genes Immun. **3**, S66-S70.

Vitte, C.; Bennetzen, J. L.; 2006 Analysis of retrotransposon structural diversity uncovers properties and propensities in angiosperm genome evolution. Proc Natl Acad Sci USA **103** 17638-17643.

Vitte, C.; Panaud, O.; 2003: Formation of Solo-LTRs through unequal homologous recombination counterbalances amplifications of LTR retrotransposons in rice *Oryza sativa* L. Mol. Biol. Evol. **20,** 528540.

Vitte, C.; Panaud, O.; Quesneville, H.; 2007: LTR retrotransposons in rice (*Oryza sativa,* L.): recent burst amplifications followed by rapid DNA loss. BMC Genomics **8**, 218

Xiao, H.; Jiang, N.; Schffner, E.; Tickinger, E. J.; Knapp, E. V.; 2008: A retrotransposon mediated gene duplication underlies morphological variation of tomato fruit. Science **139**, 1527-1530.

Xiong, Y.; Eickbush, T. H.; 1990: Origin and evolution of retroelements based on their reverse transcriptase sequences. Embo J. **9**, 3353-3362.

Received August 16, 2011