Department für angewandte Pflanzenwissenschaften und Pflanzenbiotechnologie
Institut für Garten-, Obst- und Weinbau

ANALYSIS OF TRANSPOSABLE ELEMENTS IN GRAPEVINE (*Vitis vinifera* L.)

Dissertation

zur Erlangung des akademischen Grades

Doctor rerum naturalium technicarum

an der Universität für Bodenkultur Wien

Eingereicht von

Dipl.-Ing. Andrej Benjak

Betreut von

Univ. Prof. Dr. Astrid Forneck

Wien, November 2008

ABSTRACT

The grapevine (*Vitis vinifera* ssp. *sativa*) is a widely cultivated crop that has accompanied the human culture since the domestication of this plant in the Neolithic period (8,500-4,000 BC). The cultivated grapevine derives from the wild grapevine (*Vitis vinifera* ssp. *sylvestris*), but the exact domestication events are still not clear. Sexual hybridization was the major driving force of grapevine domestication which resulted in the selection of the thousands of cultivars present today. The vegetative propagation of cultivars enabled the somatic mutations to shape the genome of individual genotypes, opening the era of clonal selection programs which resulted in the isolation of hundreds of different clones for the most established cultivars.

Transposable elements (TEs) play a major role in the dynamics of plant genomes, from their slow and long-term influence on the genome evolution to much faster phenomena like somaclonal mutations. Despite their importance, the TE content of the grapevine genome and the characteristics of grapevine's TEs were poorly analyzed so far. The sequencing of the grapevine genome made it possible to start with more detailed and genome-wide analyzes of its TEs repertoire.

This thesis is the summary of my research activities during my PhD study and presents the collection of scientific articles that are brought together in a logic unit. It starts with the introduction on grapevine, its genomics and TEs, followed by the articles: (1) Grapevine (*Vitis* ssp.): example of clonal reproduction in agricultural important plants, (2) Different DNA extraction methods can cause different AFLP profiles in grapevine (*Vitis vinifera* L.), (3) Clonal variation in Pinot noir revealed by S-SAP involving universal retrotransposon-based sequences, (4) LTR-retrotransposons of grapevine and their implementation for the IRAP and REMAP fingerprinting, (5) Genome-Wide Analysis of the "Cut-and-Paste" Transposons of Grapevine, (6) Recent amplification and impact of MITEs on the genome of grapevine (*Vitis vinifera* L.). A critical discussion on the articles is given followed by a general discussion and outlook on the topic of TEs and grapevine genomics research.

ZUSAMMENFASSUNG

Die Weinrebe (*Vitis vinifera* ssp. *sativa*) ist eine weit verbreitete Kulturpflanze, die schon in der Jungsteinzeit (8.500-4.000 v Chr.) domestiziert wurde. Die Kulturrebe stammt wahrscheinlich von der Wildrebe (*Vitis vinifera* ssp. *sylvestris*) ab. Basierend auf zufälligen Kreuzungen entstanden Tausende von Rebsorten, die heute im Anbau sind. Durch die kontinuierliche, vegetative Vermehrung von Rebsorten treten somatische Mutationen auf und gewinnen bei der Entwicklung von individuellen Genotypen (Klone) an Bedeutung, deren Selektion und Erhaltung durch Klonenselektionsprogramme ermöglicht wird.

Transposone oder "Transposable Elements" (TEs) haben maßgeblichen Anteil an der Dynamik von Pflanzengenomen bei der Evolution des Genoms und bei schnelleren Effekten wie der somaklonalen Variation. Trotz der Wichtigkeit von Transpositionen ist das Vorkommen und die Spezifika von TEs bei Reben wenig untersucht. Erst durch die Sequenzierung des Rebengenoms wurden genomeweite Analysen der TEs möglich.

Diese Dissertation ist die Zusammenfassung meiner Forschungsarbeiten, die in Form von wissenschaftlichen Artikeln, die sich in einem wissenschaftlichen Kontext befinden, präsentiert sind. Beginnend mit einer generellen Einführung in die Biologie der Rebe, der Rebengenomik und über Transposone werden die folgenden wissenschaftlichen Arbeiten dargestellt: (1) Grapevine (*Vitis* ssp.): example of clonal reproduction in agricultural important plants, (2) Different DNA extraction methods can cause different AFLP profiles in grapevine (*Vitis vinifera* L.), (3) Clonal variation in Pinot noir revealed by S-SAP involving universal retrotransposon-based sequences, (4) LTR-retrotransposons of grapevine and their implementation for the IRAP and REMAP fingerprinting, (5) Genome-Wide Analysis of the "Cut-and-Paste" Transposons of Grapevine, (6) Recent amplification and impact of MITEs on the genome of grapevine (*Vitis vinifera* L.). Abschliessend werden die Artikel kritisch diskutiert und ein Ausblick auf die genomische Forschung an TEs bei Reben gegeben.

INTRODUCTION

A BRIEF HISTORY OF GRAPEVINE

EVOLUTION OF *VITIS*

Where and when the genus *Vitis* evolved is unclear (Figure 1). It is suggested to have first appeared ~65 million years ago (de Saporta 1879, in This *et al.* 2006). The current distribution of *Vitis* species includes northern South America (the Andean highlands of Colombia and Venezuela), Central and North America, Asia and Europe. In contrast, species in the subgenus *Muscadinia* are restricted to the southeastern United States and northeastern Mexico.

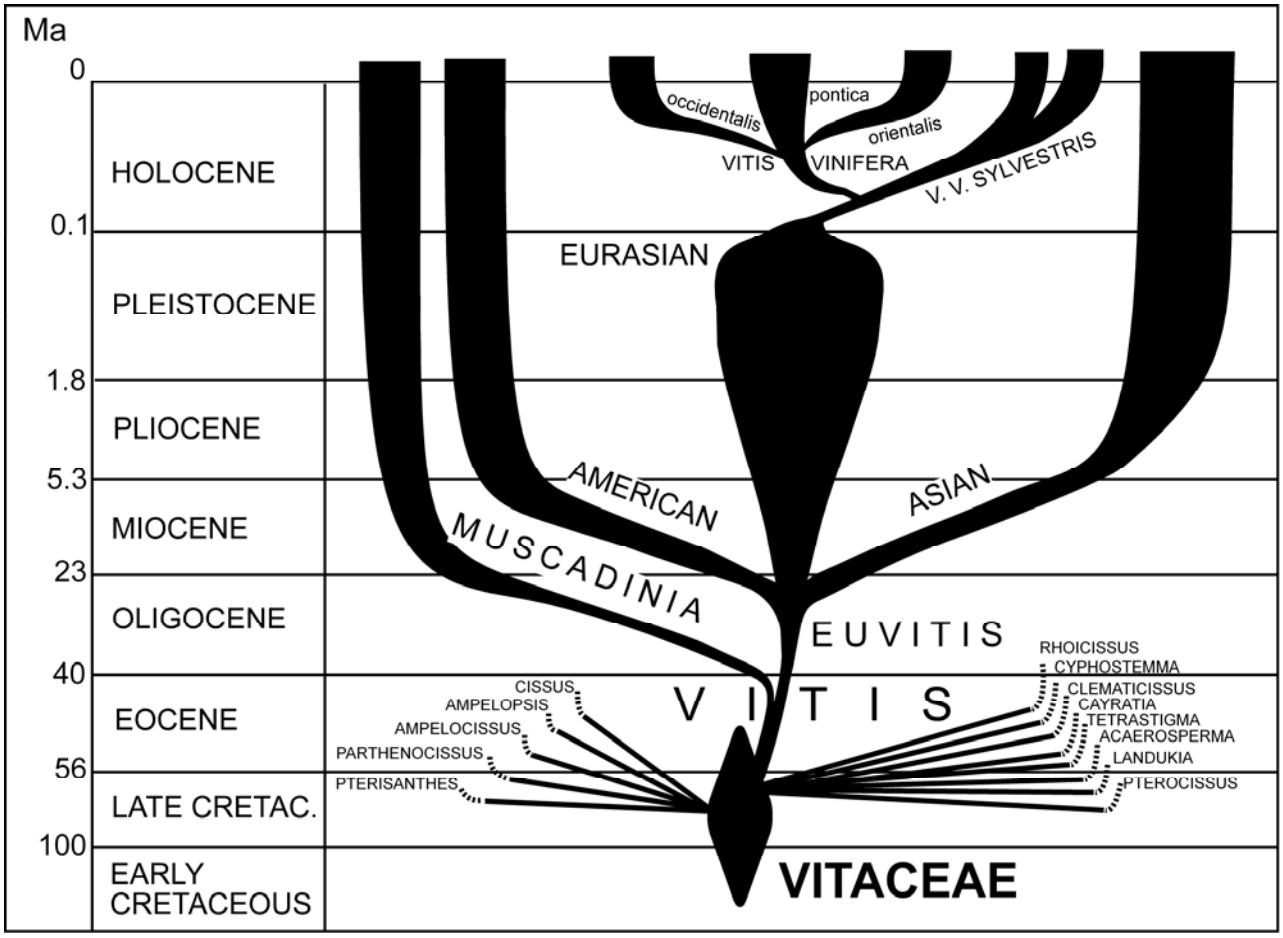

Figure 1: Evolution of the *Vitaceae* (adapted from Galet 1988).

Regardless of the manner and geographic origin of *Vitis*, the genus established its present range by the end of the last major glacial period (~8000 BC). It is believed that periodic advances and retreats of the last glacial period markedly affected the evolution of *Vitis*, notably *V. vinifera*. The alignment of the major mountain ranges in the Americas, versus Eurasia, also appears to have had an important bearing on its evolution. In the Americas and eastern China, the mountain ranges run predominantly north-south, whereas in Europe and western Asia the run principally east-west. This would have permitted North American and eastern Chinese species to move south or north, relative to movement of the ice sheets. The southward movement of grapevines in Europe and western Asia would have been largely restricted by the east-west mountain ranges (Pyrenees, Alps, Caucasus and Himalayas). This may explain the existence of only one *Vitis* sp. (*V. vinifera*) from the Atlantic coast of Europe to the western Himalayas, whereas China possesses about 30 species (Fengqin *et al.* 1990) and North and Central America some 34 species (Rogers and Rogers 1978).

DOMESTICATION OF GRAPEVINE

Cultivated grapevines (*Vitis vinifera* spp. *sativa*) are thought to have been domesticated from wild populations of *Vitis vinifera* spp. *sylvestris* (Levadoux 1956). These wild vines are dioecious plants still present in small isolated populations in Eurasia. It is very likely that the wild grape was exploited by humans in the Paleolithic era, but its domestication started later, linked to the production of wine (*ca*. 8,500-4,000 BC), even if it is unclear which process predated the other (McGovern 2003, McGovern *et al.* 1986). In Europe, evidence of wild grapevine use has been found in a Neolithic village near Paris (about 4,000 BC) (Dietsch 1996). Semidomesticated grape seed remains (2,700 BC) have been discovered in England (Jones and Legge 1987). During domestication, the biology of grapes underwent several dramatic changes to ensure greater sugar content for better fermentation, greater yield and more regular production. In this process, the changes in berry and bunch size and the change from dioecious wild plants to hermaphrodite cultivated plants were crucial. Changes in seed morphology also occurred and even if its biological significance is unknown, this trait is used in the analysis of archaeological remains to differentiate remains of wild or cultivated grape (Terral 2002). In areas where *sylvestris* vines grow in close proximity to

viticulture, the boundary between wild-types and cultivated varieties is frequently blurred by the occurrence of wild-looking escapes and by products of spontaneous hybridizations (Zohary 1996).

Major questions regarding grapevine domestication concern the number of domestication events and the geographic locations where they took place. In two recent studies (Arroyo-Garcia *et al.* 2006, Imazio *et al.* 2006) chloroplast DNA variation was analyzed at polymorphic microsatellite loci of *V. vinifera* genotypes belonging to both *sativa* and *sylvestris* subspecies. The results suggest the existence of at least two origins for grapevine cultivars, one in the Near East and a second one in the western Mediterranean region that gave rise to many of the Western European cultivars.

GENETIC VARIATION IN GRAPEVINE

GRAPEVINE CULTIVARS

Sexual reproduction and planting of seeds seems to have had an important early role in the domestication and expansion of viticulture into new regions. New genotypes are produced by sexual reproduction, either by crossing or self-fertilization. Because individual grapevine plants have highly heterozygous genotypes any progeny produced from seeds are novel combinations of parental alleles, which result in phenotypic variation and segregation of traits. The parentage studies, performed in the past decade (Aradhya *et al.* 2003, Bowers *et al.* 1999), demonstrate the importance of sexual crosses in the past for the generation of new phenotypes and the adoption and spread, by vegetative propagation, of specific genotypes with desirable characters. The *Vitis* germplasm is very variable and the wild grape germplasm is still a potential source of unique alleles for the improvement of both wine and table grapes (Aradhya, et al. 2003).

GENETIC BOTTLENECK

It is presumed that today's diversity of *V*. *vinifera* represents a scarce leftover from the diversity that existed before the spreading of the diseases from America (mainly mildews and grape phylloxera), which almost extincted the *Vitis* species. Another narrowing effect of the grapevine diversity is enhanced by economical reasons, globalization of wine markets resulting in the worldwide spreading of only few cultivars (e.g. Chardonnay, Cabernet Sauvignon, Syrah, Merlot, Pinot noir, Riesling). Other existing cultivars are less exploited, and many have only a local significance or are largely confined to germplasm collections. The number of different varieties held in germplasm collections around the world is estimated at ~10,000 (Alleweldt and Dettweiler 1994, in This, et al. 2006).

There are a number of local, almost extinct cultivars, which are not yet implemented into germplasm collections in some countries in Europe. One example is Croatia, which has diverse geographical and climatic zones and an old viticulture tradition. Unique genotypes can be expected, especially on the Croatian islands. These genotypes may have been developed within reproductive isolation and did not significantly spread out of the islands. A list of over 80 autochthonous Croatian genotypes (Pejić *et al.* 2000) exists, which is now being implemented in germplasm collections in Croatia. Some genotypes are represented by only few old individuals (Pejić and Maletić, personal communication) and is possible that undiscovered and rare genotypes exist which could face extinction soon.

Another way of impoverment of the European *Vitis* germplasm comes through the slow extinction of the *sylvestris* population which is caused mainly as a consequence of the introduction of pathogens from North America and of the destruction of its habitat. Attempts to reintroduce wild grapevine into its natural habitat were made in the last decade, but no satisfactory results were achieved (Arnold *et al.* 2005).

CLONES AND CLONAL VARIATION

A clone is defined as an individual that descended from a single common ancestor by mitosis. A group of clones originating from the same ancestor (monozygotic) is defined to be genetically identical except for the effect of mutations (Forneck 2005).

Clonal propagation has sustained cultivar identity but restricted the improvement of these genotypes to strategies that do not involve conventional breeding. Clones are selected within a cultivar which exhibit phenotypic differences. Clonal selection is based on the genetic variation within cultivars. Possible explanations of clonal variability are the following: polyclonal origin of cultivars, pathogen infections, somatic mutations, epigenetics and chimeras.

POLYCLONAL cultivars derive from more than one seedling of the same parents which are phenotypically very similar. From a genetic point of view, “polyclones” would be considered as different cultivars because each group of clones in a polyclonal population derives from

different hybridization events. Still, because of the phenotypical similarity and practical reasons, “polyclones” are often considered as members of one cultivar.

PATHOGEN INFECTIONS, especially virus diseases contribute to increase the phenotypical variation within grapevine cultivars. Therefore, these are not genomic clonal variation, but external influences. The problem arises in cases when a virus does not provoke any obvious symptoms, but rather changes the phenotype slightly. This may result in the pre-selection of a number of false candidates during clonal selection. To overcome this problem, modern techniques (ELISA tests or PCR based methods) are available to detect viruses in grapevine and sanitary inspection is one of the first steps in clonal selection.

EPIGENETICS are features that are stable over rounds of cell division but do not involve changes in the underlying DNA sequence. The molecular basis of epigenetics is complex. It involves modifications of the activation of certain genes, but not the basic structure of DNA. Major mechanisms are DNA methylation, chromatin remodeling and RNA interference (Henderson and Jacobsen 2007). Various aspects of epigenetic control in plants can be influenced by stress (Boyko and Kovalchuk 2008), which is interesting for cultivated grapevine considering the extensive pruning that these plants constantly undergo, exposure to various climatic conditions and in some cases application of tissue culture for propagation.

SOMATIC MUTATIONS in term of clonal variability are differences in the nucleotide sequence of the genetic material among clones. Mutations are relatively rare and are always local in multicellular organisms, which means that they occur occasionally in individual cells. If the affected cells are meristematic, the mutations will be present in all organs and tissues that develop from these cells. Mutations can be classified according to different categories. By effect on structure, mutations can be small-scale (point mutations, short insertions, deletions etc) and large-scale (amplifications, large deletions, chromosomal rearrangements etc). By aspect of phenotype altering, mutations can be silent or functional. Silent mutations do not affect any aspect of genomic activity while functional mutations cause changes in the genomic system which can affect any aspect of gene regulation and control, change enzymatic pathways (biochemical mutations) or lead to differences in the outward appearance of an individual (morphological mutations). It has to be noted that these

classifications are rather theoretical than practical. This is because, given the complexity of the genomic system, it is difficult to predict whether a certain mutation will have any functional impact on the genome or not. In practice, clonal selection of grapevine relies only on those mutations that change the phenotypic properties of interest. It is also possible to imagine that certain phenotypic changes occur only after an accumulation of specific mutations, each of which alone does not have a significant impact on the genome. Therefore, intraclonal genomic variability is always present, even though subtle, challenging the goal of molecular breeding which is to identify and locate all genomic factors that control a specific trait.

CHIMERAS (in botany) are single organisms composed of two genetically different types of tissue. The difference is often due to somatic mutation during ordinary cell division. Chimeric plants have specific phenotypes due to the interaction of genetically distinct cell layers. This phenomenon is observed, for example, in Pinot Meunier (Franks *et al.* 2002). Because of the higher somatic variation of a single plant, chimeras complicate fingerprinting and genetic studies. This is especially true for grapevine, firstly because fingerprint profiles are currently one of the main cultivar identifiers which are also widely used for parentage studies, and secondly because grapevine is vegetatively propagated (for centuries for many cultivars) which increases the chances for development of chimeric plants. Chimeric plants with morphological mutations are usually easy to spot because cells from the L2 layer occasionally (and locally) spread into the L1 layer leading to a mosaic phenotype. There is no doubt that many grapevine chimeras exist caused by “minor” biochemical mutations, while chimeras with silent mutations are probably even more common, just very difficult to characterize.

METHODS FOR DETECTING SOMATIC MUTATIONS

Up to now, there are no methods able to identify all mutations in a given genome. There are several fingerprinting methods which are used to identify genomic interclonal variation, the most efficient so far being the SSR, AFLP, SNP, S-SAP and derivates of these methods (see Forneck 2005 for review on their application). But these techniques are only able to detect

polymorphism to a limited extend and relatively quantify them. The problem of identifying all genomic differences between two close related genomes might soon be overcome by ever cheaper and more efficient whole-genome-sequencing services. Individual plants will be sequenced and close related genomes compared, a practice which is currently performed only in few cutting-edge scientific projects, but might be done routinely very soon.

On the other hand, detecting and targeting genomic polymorphism is only the first step to characterize mutations. How mutations could affect a genome is the most challenging question in genomics today. Given the complexity of this problem, the only way to resolve it is to dissect the problem it into smaller components and closely focus on each of them. One of the major mutagenic factors and important genomic regulators are transposable elements.

TRANSPOSABLE ELEMENTS

DEFINITION AND CLASSIFICATION

Transposable elements (TEs) are segments of DNA that have the ability to move and/or replicate within genomes. They were discovered by the Nobel Prize winner Barbara McClintock and have been found to be ubiquitous in most living organisms (McClintock 1956). They are present in copy numbers ranging from few up to millions per genome. TEs can represent a major fraction of the genome, especially in plants (Kidwell 2002).

TEs are classified into two major classes depending on the type of transposition intermediate: class I TEs, which have an RNA transposition intermediate, and class II TEs, which have an DNA transposition intermediate (Finnegan 1989). There are currently few suggested nomenclatures for TEs (Kapitonov and Jurka 2008, Wicker *et al.* 2007) but they all relay on a common, more or less identical classification.

Class I elements or retrotransposons transpose in a copy-paste fashion through reverse-transcription of a RNA intermediate, which is encoded by the element. Retrotransposons can be divided into two principal groups, the LTR and the non-LTR retrotransposons. LTR retrotransposons have direct long terminal repeats (LTRs) of variable length (from 100 bp to several kbs) that flank the internal coding region. Both groups encode a number of proteins in two major genes, *gag* and *pol*, that are synthesized as a polyprotein, which is then cleaved into multiple functional peptides by an element-encoded protease. *Gag* encodes structural proteins important for the packaging of the retrotransposon RNA while the *pol* gene encodes the enzymes needed for the retrotransposition (protease, integrase, reverse transcriptase and RNaseH). The two major groups Ty1-*copia* and Ty3-*gypsy* differ in the position of integrase within the encoded polyprotein. LTR retrotransposons are especially abundant in plants with large genomes. They comprise 50-80 % of the maize (Meyers *et al.* 2001, Sanmiguel and Bennetzen 1998) and barely genome (Vicient *et al.* 1999). In plants with smaller genomes LTR retrotransposons constitute a smaller percentage of the genome,

like in rice <10 % (Mao *et al.* 2000) and *Arabidopsis* ~5 % (Kapitonov and Jurka 1999). Non-LTR retrotransposons (LINEs and SINEs) lack LTRs and are transcribed from an internal promoter. LINEs (long interspersed nuclear elements) have *gag* and *pol* genes and are common in plants while SINEs (short interspersed nuclear elements) lack any coding capacity and are more common in animals (Schmidt 1999).

Class II elements or DNA transposons transpose mostly in a "cut and paste" fashion through a DNA intermediate. They encode for a transposase that recognizes its terminal inverted repeats (TIRs), excises the element and inserts it elsewhere in the genome, generating short target site duplications flanking the element. Elements that have lost the ability to encode a transposase (non-autonomous elements) require the transposase of an autonomous element of its family to transpose. The presence of conserved motives within transposases, as well as sequence and length similarities in the TIRs and in the target site duplications generated upon insertion, allow classifying eukaryotic DNA transposons in 5 to 7 different superfamilies (Feschotte *et al.* 2002b, Robertson 2002), the most important for plants being the *Mutator*, *CACTA*, *hAT*, *PIF* and *Mariner* superfamily. Miniature Inverted repeat TEs (MITEs) are a particular type of defective class II elements characterized by their small size (from 150 to 600 bp) and high copy numbers (Moreno-Vazquez *et al.* 2005). MITEs are distinguished from other non-autonomous class II transposons by the high uniformity of their copies and in some cases with the potential to form secondary DNA structures. Although they are mobilized by transposases and encoded by their related autonomous elements (Feschotte *et al.* 2005, Loot *et al.* 2006), the mechanism by which these elements are amplified remains unknown. Additional exceptions from DNA transposons with standard structure and transposition mechanism are *Arnold* and *Vandal*, which have no TIRs (Kapitonov and Jurka 1999), *Helitrons* which transpose by rolling-circle replication (Kapitonov and Jurka 2001) and *Polintons* which encode up to 10 different proteins (Kapitonov and Jurka 2006).

PHYSICAL REMODELING OF GENOMES

The main property of TEs is to transpose, which means causing insertional polymorphisms in genomes. Class II elements additionally cause deletions through their excisions. Transpositions *per se* are mutations, some of which might alter the phenotype, especially when genes and regulatory factors are affected by this process. The most visible mutation that TEs can cause is gene knockdown by inserting into or close to it. For example, the insertion of the LTR-retrotransposon *Gret1* into the promoter of the *Myb*-related gene that regulates anthocyanin biosynthesis causes a white-berried grapevine phenotype (Kobayashi *et al.* 2004). This mutation is present in most white grapevine cultivars (Kobayashi, et al. 2004, This *et al.* 2007).

Consequences of transposition activities go further than simple insertion/deletion. One is the increase of genome size by the amplification of retrotransposons which might have consequences in genome stability. Transposition of retrotransposons includes an amplification step via RNA intermediates by the reverse transcriptase. In maize, LTR-retrotransposons make up over 70 % of the nuclear genome (Sanmiguel and Bennetzen 1998) and they are major contributors to all other large plant genomes. On the other hand, class II TEs in general do not copy themselves and are instead amplified indirectly by the host DNA repair machinery. The gap in the DNA caused by the double strand DNA cleavage of the transposase and excision of the TE is replaced with the homologous chromosome restoring the TE, while the original TE transposes somewhere else. Exceptions are MITEs and *Helitrons* which are usually present in high copy number in genomes (Kapitonov and Jurka 2001, Naito *et al.* 2006). To counter the effect of "genomic obesity" caused by TEs, genomes adopt strategies like unequal homologous recombination and illegitimate recombination to generate abundant small deletions that can attenuate or reverse plant genome growth (Bennetzen *et al.* 2005). These mechanisms could also cause mutations and chromosome rearrangements.

The first property of TEs to be recognized was their ability to break chromosomes (McClintock 1947). This may be caused when two identical TEs are inserted into each other

in inverse orientation. Dicentric chromosomes are then formed from a transposition reaction involving TE ends in sister chromatids (English *et al.* 1995). Macrotranspositions and complex chromosome reatangements can occur when TEs are present in specific arrangements in the genome (Huang and Dooner 2008).

Class II TEs can capture host genome sequences and mobilize and amplify them together with their own sequences in a process known as transduplication. This is particularly common for the *Mutator*-like elements (MULEs) (Hoen *et al.* 2006, Holligan *et al.* 2006, Jiang *et al.* 2004, Le *et al.* 2000, van Leeuwen *et al.* 2007, Yu *et al.* 2000) and the *CACTAs* (Kawasaki and Nitasaka 2004, Zabala and Vodkin 2005) and was recently found also in *PIFs* (Benjak *et al.* 2008) and MITEs (Benjak *et al,* in prep). Although most of these captured gene fragments seem to be non-functional pseudogenes (Hoen, et al. 2006), it has been recently reported that, in some cases, transduplicated exons could be incorporated into host transcripts by alternative splicing giving rise to new host proteins (Zabala and Vodkin 2007). Even having lost their coding capacity, transduplicated sequences may undergo transcription and have a regulatory function (Hoen, et al. 2006). In the case of retrotransposons, this phenomenon is known as transduction and involves readthrough transcription from a retrotransposon promoter into adjacent host gene sequences and the incorporation of the host gene into the transposon sequence during reverse transcription (Bureau *et al.* 1994). Such events can lead to formation of novel hybrid open reading frames (Elrouby and Bureau 2001) and have an impact of gene evolution.

GENE REGULATION

In addition to physical remodeling of genomes, TEs can influence gene regulation as well. For example, *Ds1* element in exon 9 of the maize *waxy* gene causes several alternative splicing of its pre-mRNA (Wessler 1991). Moreover, TEs are tightly associated with transcriptional and post-transcriptional gene silencing. DNA methylation is used (especially in plants) to inhibit the transcription of TEs but is also linked to stress response (Rabinowicz *et al.* 2003). Methylated TEs can affect the transcription of nearby genes (Weil and Martienssen 2008) because DNA methylation directly affects histone modification and

chromatin remodeling, which in the end affects transcription. This cascade of processes places TEs as important factors in epigenetic control (Costa 2008, Henderson and Jacobsen 2007).

TEs have an important role as source of micro-RNAs (miRNAs) and small interfering RNAs (siRNA). miRNAs are a class of short, 22nt non-coding RNA (ncRNA) that function as posttranscriptional regulators of gene expression (Bartel 2004). Mature miRNAs are processed from longer RNA sequences that form local stemloop (hairpin) structures. The mature miRNA sequence binds to partially complementary target sites in the 3' untranslated regions (UTRs) of messenger RNAs (mRNAs) and regulates expression through a process of mRNA degradation and/or translational repression (Bartel 2004). siRNAs are closely related to miRNAs in terms of both biogenesis and regulatory function. The difference is that siRNAs are generated from long dsRNA precursors, which can be either endogenous or exogenous transcripts, whereas mature miRNAs are processed from shorter endogenous transcripts that form local hairpin structures (Buchon and Vaury 2006). One previously recognized distinction between these two classes of regulatory RNA is the fact that miRNAs are generally found in unique genomic loci, such as intergenic regions, while siRNAs originate from within already characterized sequences such as genes and transposable elements (TEs) (Matzke *et al.* 2000, Slotkin *et al.* 2005, Vastenhouw and Plasterk 2004). However, recent reports indicated that a number of mammalian miRNAs are in fact derived from TEs (Piriyapongsa and Jordan 2007, Smalheiser and Torvik 2005). TEs may also provide an evolutionary connection between siRNAs and miRNAs. In fact, some believe that the widespread RNA-silencing pathways originated from an ancestral immune system aimed at transposable elements, viruses and other intracellular invaders. RNA-silencing not only affects transcripts, but can also target DNA to induce methylation and is involved in the maintenance of heterochromatin and the silencing of TEs (Madlung and Comai 2004). The abundance of TEs, their repetitive nature and specific structures provide a natural mechanism for the generation of multiple RNA interfering pathways that affect gene expression (Thornburg *et al.* 2006).

DOMESTICATION

From insertional mutagenesis and gene regulation TEs go even further in the interaction with their host genomes in a process called transposon domestication. TE domestication means that the host genome is using some TEs as regular genes. Such TEs are under the same selective pressure as genes and undergo modifications throughout their evolution, such as losing their ability to transpose. This is achieved by deletion of their TIRs or LTRs (depending on the class or TE) and modifications on the proteins that they encode (usually only the binding capacity is maintained). Even though such genes can be referred as "TE-derived genes", their sequence homology to related TEs can be so high that it becomes difficult to distinguish them from other TEs. Therefore the term "domesticated TEs" is widely used. If not properly analyzed, domesticated TEs will be annotated as regular TEs in homology based annotations. Therefore, closer attention is given to candidates for domesticated TEs, which are single copy elements that lack some structural features common for their class. Domesticated TEs were found for all major TE groups, and for some of them their homologues are found in different species which means that the strategy for the genome to use TEs as source of novel genes is evolutionary very old.

Examples of plant domesticated transposases are the *Arabidopsis* transcription factors *FAR1* and *FHY3*, derived from MULE transposases (Hudson *et al.* 2003), or *DAYSLEEPER*, a gene essential for *Arabidopsis* development which probably encodes a transcription factor derived from a *hAT* transposase (Bundock and Hooykaas 2005). Other domesticated transposons which function is not known are the *MUSTANG* and the *Gary* elements, the former originated from MULEs and the later from *hAT* transposons (Cowan *et al.* 2005, Muehlbauer *et al.* 2006).

Cases of Class II TEs domestication were found as well, such in the case of the *AtCopeg1* gene which evolved from *AtCopia95* family of retrotransposons in *Arabidopsis* (Duan *et al.* 2008). Another example is a new group of elements, aptly named diversity-generating retroelements (DGRs), that have been coopted by temperate bacteriophages to bypass the defenses of their bacterial host *Bordetella* (Feschotte and Pritham 2006).

Not only coding regions of TEs are being domesticated. For example, MITEs can give rise to novel miRNA genes (Piriyapongsa and Jordan 2007). MITEs can serve as binding sites for their related and domesticated elements, like in the case of SETMAR, a primate-specific gene that arose by fusion of a histone methyltransferase SET domain with a mariner-like transposase (Liu *et al.* 2007, Miskey *et al.* 2007).

CONCLUSION ON TES

Transposable elements are known to be major contributors to genome variability and, in particular, to somatic mutations. This is the reason why I have chosen to focus my research on TEs. Several questions have been brought at the start of my work, like what is the TE content of grapevine? Are there still active elements? Do TEs affect genes in grapevine? Can TEs be used for genetical studies in grapevine? And so on. Some of these questions have been answered and the research behind the answers is given in the following chapter.

ARTICLES

The following section presents the scientific articles that were done during my PhD study. Some articles that were still not submitted for publication are in form of draft manuscripts. The first article, “Grapevine (*Vitis* ssp.): example of clonal reproduction in agricultural important plants” is a part of the book called Lost Sex: The Biology of Parthenogenetic Organisms (in press). This review article is a very good introduction for the following section because it deals with the vegetative propagation of grapevine, grapevine clones and clonal variability.

The second article, “Different DNA extraction methods can cause different AFLP profiles in grapevine (*Vitis vinifera* L.)” is a methodological article that stresses the issue of DNA quality for DNA analyzes. This article is useful to a wider spectrum of research involving DNA manipulation.

The following two articles (“Clonal variation in Pinot noir revealed by S-SAP involving universal retrotransposon-based sequences” and “LTR-retrotransposons of grapevine and their implementation for the IRAP and REMAP fingerprinting”) present the implementation of different types of fingerprinting methods for grapevine using LTR-retrotransposon sequences. Preliminary results on genome-wide analysis of LTR-retrotransposons are given as well.

The last two articles (“Genome-Wide Analysis of the “Cut-and-Paste” Transposons of Grapevine” and “Recent amplification and impact of MITEs on the genome of grapevine (*Vitis vinifera* L.)”) present a genome-wide analysis of class II transposons of grapevine. Focus is given to the characteristic properties of analyzed TEs and their contribution to the evolution and variability of the *Vitis* genome.

ARTICLE 1

Grapevine (*Vitis* ssp.): example of clonal reproduction in agricultural important plants

Astrid Forneck[1], Andrej Benjak[1], Ernst Rühl[2]

[1]Department of Appl. Plant Sciences and Plant Biotechnology, University of Natural Resources and Appl. Life Sciences, Vienna, Peter-Jordan Str. 82, A-1190 Vienna, Austria.

[2]Geisenheim Research Centre, Section for Grapevine Breeding and Grafting, von Lade-Str. 1, D-65366 Geisenheim, Germany.

Corresponding author: Astrid Forneck email: astrid.forneck@boku.ac.at

In: Lost Sex: The Biology of Parthenogenetic Organisms, (eds. Isa Schön & Koen Martens), Springer-Verlag, Berlin (in press).

Inclusion permission granted by publisher.

29

Grapevine (*Vitis* ssp.): example of clonal reproduction in agricultural important plants

Astrid Forneck[1], Andrej Benjak[1], Ernst Rühl[2]

[1]Department of Appl. Plant Sciences and Plant Biotechnology, University of Natural Resources and Appl. Life Sciences, Vienna, Peter-Jordan Str. 82, A-1190 Vienna, Austria.

[2]Geisenheim Research Centre, Section for Grapevine Breeding and Grafting, von Lade-Str. 1, D-65366 Geisenheim, Germany.

Corresponding author: Astrid Forneck email: astrid.forneck@boku.ac.at

Abstract

This review approaches the concept of clonality (asexual reproduction) and its implications for phenotypic and genetic stability. Grapevine cultivars (*Vitis vinifera* L.) are composed of clones showing homogeneous ampelographic characteristics with minor differences. The concept of clonal selection (through vegetative propagation) insinuates very low genetic variation within a "population of genotypes identical to the ancient progenitor except of mutations". Yet, the genetic variation that in cultivated grapevine clones accumulates is higher than expected. This variation is further increased by numerous mechanisms of an asexual life strategy to enhance variation and to provide an open system for adaptation and selection processes. This chapter also provides insight into the clonal selection of grapevine exemplifying the cultivar Pinot noir (*V. vinifera* L.). The impact of clonal propagation of this agricultural important crop will be discussed.

29.1 Introduction

Grapevine (*Vitis vinifera* L.), economically one of the most important crop plants, comprises multiple varieties and clones. The identification and characterization of grape cultivars (varieties) has always been an intrinsic concern for agriculture as well as for breeding research programs. Traditionally, morphological characteristics of the plants have been implemented for distinguishing grapevine cultivars and often resulted in insufficient or even unsuccessful differentiation. Grapevine is commercially propagated vegetatively and cultivars existing today result from the selection of advanced genotypes of ancient origin mostly generated by spontaneous crosses centuries ago (Mullins and Meredith 1989). Each ancient cultivar expresses distinct phenotypes, resulting in sets of morphologically diverging clones. These clones are spread worldwide adjusting to different environments and cultivation techniques. Currently almost 16.000 prime named grapevine cultivars are listed in the International *Vitis* Variety Catalogue (http://www.genres.de).

Grapevine domestication from *Vitis sylvestris* Gmel. or *Vitis caucasica* Vav. dates back to the Neolithic period (Negrul 1946, Levadoux 1956, Ambrosi and Becker 1978). Apart from their various usages as fresh fruit, dried fruit, jam, wine or vinegar, the simple vegetative multiplication was a key reason for early domestication. Vegetative propagation of grapevines is straightforward: the long and flexible wooden canes are used for “natural” layering and single, selected dormant cuttings can be easily rooted. In this way, clonal selection may have additionally triggered grapevine domestication, since promising phenotypes were multiplied, spread according to human transportation and conserved over centuries. Vegetative propagation has consequently been advantageous and used from the very beginning of grapevine domestication (Billiard 1913) and many fruit species as well (Zohary and Spiegel-Roy 1975). But sexual reproduction was not completely evaded either, though generative propagation of grapevine is difficult: seeds germinate erratically and plants grown from seeds vary extremely, due to high levels of heterozygosity (Bowers et al. 1999). Parentage studies show that natural crosses must have happened. Table grapes were eaten and the seeds spat out or exuded, wine grapes were pressed and the pomace dumped in the vicinity of the winery, leading –though rarely- to superior varieties. Some varieties appear to be immediate selections from wild types, e.g. the variety Traminer, (Regner 1999) while others are crosses between existing cultured varieties, e.g. Cabernet

Sauvignon, a cross between Sauvignon blanc and Cabernet franc (Regner et al. 1998, Bowers and Meredith, 1997) or crosses between wild types and cultured varieties, e.g. Riesling, a cross between Gouais and most likely a Traminer *V. silvestris* hybrid (Regner 1999). In many cases, parent varieties were at their time important varieties, but have virtually disappeared from modern viticulture, like the variety Gouais also called Heunisch in Germany, which is a parent of more than 70 different cultivars (Boursiquot et al., 2004), e.g. Chardonnay or Gamay (Bowers et al., 1999).

For grapevine clones the concept of individuality is straightforward and relies on propagation records and morphological features leading to the breeding concept of clonal selection. The first description of the need of clonal selection and useful methods are found in Roman sources as indicated by Columella (60 b. Chr.), emphasizing regular visual evaluations and positive mass selection according to quality-related traits (yield, fruit set). Since grapevine is a high priced crop with significant viticultural research in describing and analyzing the phenotypes has been performed since the 19th century. Subsequently with the onset of genomic research *Vitis* ssp. has been focus of many studies using elaborate tissue culture, transformation and molecular genetic techniques. In general, investigations on clonal variation within grapevine cultivars have shown that the degree of detected genetic divergence usually depends on the applied marker system and on the scope and type of plant samples (Forneck 2005). The retrotransposon-based marker systems SSAP or ISTR have shown higher levels of polymorphism (Labra et al. 2004, Sensi et al. 1996) than the standard AFLP or SSR techniques (Blaich et al. 2007, Konradi et al. 2008, Regner et al. 2000).

In this chapter we intend to explain the importance of asexual as in vegetative propagation and clones for an agricultural crop leading to the breeding strategy of clonal selection. We give a short review on the successes of clonal selection and discuss the mechanisms behind the clonal variation in grapevine introducing the variety Pinot noir as a well analysed example, revealing high genetic similarity among individual clones and also indicating their origin in asexual reproduction (Ye et al. 1998, Regner et al. 2000). Compared to other grapevine cultivars Pinot noir clones are characterised by high phenotypic diversity, originated by spontaneously occurring mutations of several kinds. Several pale coloured mutants have emerged from the red grape Pinot noir. Pinot gris has been identified as periclinal chimera resulting from a somatic mutation at the berry colour locus (Walker et al.

2006). The white-skinned Pinot blanc is considered to have also arisen from Pinot noir. The insertion of a retrotransposon into one as well as the deletion of the other allele of the *VvmybA1* gene has blocked the production of anthocyanin in the white grape (Yakushiji et al. 2006). To further contribute to the understanding of sexual propagation in grapevine, we add a short summary of the currents status quo on grapevine genetics and breeding to highlight that sex has not been lost in grapevine and conclude with the advantages of grapevine clones for viticulture.

29.2 Clonality in Grapevine

Clonal selection – the art of bringing clonal variation to the fields

A grapevine clone is the vegetative progeny of a single plant. In the absence of mutations all descendants of a clone have identical phenotypes and genotypes. Modern clonal selection started in 1876 when a wine grower did research on quality traits (yield) of single Grapevine plants over a period of 20 years. The result of this work contributed to the first registered "grapevine clone" of the variety *V. vinifera* cv. Silvaner with an average yield of 6.637 kg per vine (Froelich 1900). This success of Froelich's approach resulted in numerous activities in clonal selection of grapevines, first in Germany, then in most vine growing countries. Today, clonal propagation material is available from almost all important varieties and used world wide. Currently in Germany ca. 600 clones are registered (Becher 2007) and in France more than 1000 (Boidron et al., 1997). Phenotypic differences among clones of the numerous grapevine cultivars have been reported by many authors (e.g. Sievers, 1971, Silvestroni et al., 1995, Boidron et al., 1997, Rühl et al**.** 2000). Long-living grapevine is prone to adapt to environmental and pathogen effect, thus leading to both phenotypic and genotypic alterations, that may mimic clonal variation. Clonal variation here is defined as changes in genomes other than sexually derived, that will be transmitted asexually to the descendants.

Virus infection can significantly alter the performance of vines. Inoculating *V*. *vinifera* cv. Albana and cv. Trebbiano Romagnolo vines with Grapevine Fanleaf or Grapevine Leafroll virus, reduced yield - depending on the virus type - by up to 72.9% and 80.4%, respectively Credi and Babini (1997). Consequently, virus elimination can affect vine performance

significantly (e.g. Mannini et al., 1994, 1998) and virus freedom is consequently an essential prerequisite for the production of grapevine clones to be used as propagation material (Walter and Martelli 1996, 1997).

For clonal selection phenotypic variation within a cultivar is to be clearly identified. While qualitative traits (e.g. number of bunches per shoot or bunch architecture) may be recognised on a single vine, quantitative traits (e.g. yield, sugar production, acidity) are only noticeable in larger plantings in experimental designs set up for clonal selection. The aims of clonal selection largely depend on the cultivar and its use.

Clonal selection an example from V. vinifera cv. Pinot noir

To illustrate clonal variation in grapevine, we exemplify a study on *V. vinifera* cv. Pinot noir. Forty-four Pinot noir clones were virus tested in spring 1988, grafted on Börner rootstocks and planted in a field trial in spring of 1989 in a fertile sandy loam at Geisenheim, Germany. The results are means of six years. Large clonal variation was found in three phenotypic traits: yield, acid content (titratable acidity) and susceptibility to botrytis bunch rot, the major disease of ripening berries caused by *Botrytis cinerea*. The sugar content, measured as total soluble solids, showed only small variation between clones (Figure 1). One clone had an average yield of less than 700 g m^{-2} **,** while one produced more than 1600 g m^{-2}. Titratable acidity of different clones ranged from less than 9.5 g L^{-1} to more than 13.5 g L^{-1} and bunch rot susceptibility varied between less than 2 % to more than 26 % botrytis infected berries. Titratable acidity and bunch rot incident did not resemble a normal distribution. Plotting berry sugar content of a clone as a function of its corresponding yield revealed a trend to lower sugar levels with increasing yields (Figure 2a); clones with higher yield have lower sugar levels and *vice versa*. Clones may be grouped by a system published by Oustric (1994). A vertical line at the average yield and a horizontal at the mean total soluble solids value divide the graph in 4 quadrants. Quadrant A holds clones with generally low yield and high sugar level, B clones with both a high yield and high sugar level, C high yielding clones with low sugar level and D clones with both low yield and low sugar level. A-clones are well suited for the production of premium wines, B-clones depending on the

cropping level may be used for both premium and bulk wine production and C-clones mostly for bulk wines. Variation becomes even more obvious, if bunch rot incidence is plotted as a function of titratable acidity (Figure 2b). Four major phenotypically diverging groups of clones emerge: clones with compact bunches, high risk of bunch rot and high acidity; clones with upright growing shoots, average incidence of botrytis and average acidity; clones with loose clusters, low acidity and low botrytis risk and clones with small berries, high acidity and low incidence of botrytis (Lindner et al., 1999, Rühl et al., 2000). With the choice of the right clone, a grower largely determines the plant performance and fruit quality in the vineyards.

29.3 Sources of clonal variation in grapevine explored in viticulture

Knowledge about grapevine genetics is still scarce and new discoveries will allow to better understand how naturally occuring mutations influence the phenotype of different cultivars and their clones and to adequately modify some of their genomic properties for a more successful breeding. The grapevine genome consists of various confirmed sources of genetic variation relying on mutation events. Along with the mutability in random soma and germinal cells, there is an interacting driving force among tissues resulting in chimeric structures. There is abundant evidence for the occurrence of somatic mutations in plants. Within-individual variability for polygenic and cytogenetic traits has been documented by several authors (e.g. Klekowski and Godrey 1989). In addition, molecular genetic variation has been found among naturally occurring clones in several plant species (Tuskan et al. 1996; Capossela et al. 1992), though little is known about mutation rates during somatic development of plants (Gill et al. 1995). One explanation for variation within the progeny is the occurrence of spontaneous mutations (Forneck 2005) which may be traced by genetic fingerprinting techniques.

Clonal variation by random mutation

The objective of genetically fingerprinting grapevine clones is to confirm genetic similarities and to search and identify reproducible sequence mutations. Clonal fingerprinting based on DNA sequence alterations has been performed with various PCR-assisted marker systems and bases on the assumption that a distinct individual DNA exists in each individual plant. Experimental evidence shows that this assumption does not generally hold (Blaich et al. 2007). Chimerism, tissue-specific and time-specific methylation, stress-related dynamic transposition events (Benjak et al. 2008) exemplify the multitude of processes resulting in genomic expansion. As to date we cannot denounce the existence of rapidly micro-evolving genomes reflecting dynamic sequence mutations that are not soundly transmitted in subsequent vegetation cycles, but can be traced by highly sensitive genetic marker techniques. Recently, two articles describing the sequence of the Vitis genome (*V. vinifera* cv. Pinot noir) have been published (Jaillon et al. 2007; Velasco et al. 2007) and a draft sequence of grapevine genome has been made available, opening the possibility for a genome-wide bioinformatical analysis for clonal variation. "Measuring" clonal variation among grapevine clones of a given cultivar remains difficult, since the original motherplant and its genotype are rarely known or extinct. One way to approach the nature of mutation events among agronomically cultivated grapevine clone may be the identification of a "most common clonal genotype". This has been approached by Hocquigny et al. (2004) analysing 145 accessions, belonging to five Pinot cultivars (*V. vinifera* L.) at 50 loci. A Pinot "genotype I" has been proposed due to the facts that (1) 65 % of all samples shared this genotype, (2) the remaining variant clones shared a minimum of 95 % of all loci and (3) most of the loci showed fixed heterozygosity. This study postulates that genotype I is likely the most common ancestor of five Pinot cultivars, which means these were generated by asexual propagation from a single unique zygote. A study analysing the clonal variation among seventy Pinot noir (*V. vinifera* L.) clones deriving from a single cultivar but displaying various phenotypes (cluster architecture, maturity, canopy growth) implemented 178 AFLP-markers in a replicated, stringent design. A "most common" genotype, comprised eighteen (25.7 %) identically fingerprinted clones. The biggest group of 48 clones (68.6 %) was analyzed within the range of 99 % genetic similarity compared to the main identical group. A group of 24 clones could be situated further than 1 % differentiation: among 99.1 – 94.0 % genetic

similarity. The identification of clonal variation depends on the sample size and the molecular marker system selected. Furthermore, the selected clone samples play a major role. Given the possible existence of one common zygote for all Pinot clones (approximately 500) one would have to identify the major selection lines prior to search for random mutations, since otherwise these may mimic the genetic variation exiting in world wide Pinot selections.

Clonal variation by transposition

As earlier studies on other organisms (reviewed in Bennetzen 2000) as well as grapevine itself showed (Kobayashi, et al. 2004, Verries *et al.* 2000), the modifications of the genome induced by transposable elements are one key to our understanding of grapevine genetics. Transposable elements (TEs) are DNA segments possessing the ability to move or multiply within genomes, thereby generating self-copies interspersed with non-repetitive DNA (reviewed in Feschotte *et al.* 2002a). Many of them encode protein(s) required for their mobility and solely use the host cellular machinery for their transcription and translation. These are called autonomous elements. On the other hand, the mobility of non-autonomous elements relies on proteins encoded by related autonomous TEs. TEs are classified into two classes based on their mechanism of transposition (Finnegan 1989): the class I elements, also called retrotransposons, use a RNA intermediate and a reverse transcriptase for their transposition, whereas the class II elements, or DNA transposons, use a DNA intermediate and a transposase. Retrotransposons are divided into two principal groups, the LTR (Long Terminal Repeat) and the non-LTR retrotransposons. DNA transposons have Terminal Inverted Repeats (TIRs) flanking the gene for a transposase (in the autonomous elements). Eukaryotic DNA transposons are classified into 5 to 7 different superfamilies (e.g. Feschotte, et al. 2002b, Robertson 2002). MITEs are sepcial group of class II elements which are characterized by their small sequence size (100-500 bp) and the usual preference for insertion in genic regions.

RAPD amplification products have provided the first indication for the presence of retroelements and remnants thereof in the grapevine genome. Cloned repetitive sequences

showed high similarities to retrotransposons of higher plants, and found to be dispersed throughout the genome (Böhm and Zyprian 1998). So far, only few TEs (LTR-retrotransposons) were described in grapevine, namely *Tvv1* (Pelsy and Merdinoglu 2002), *Vine-1* (Verries, et al. 2000), and *Gret1* (Kobayashi, et al. 2004). The last two were found to be inserted in the *Adhr* and the *VvmybA1* genes respectively, confirming that TEs have the potential to alter genes in grapevine. Recent studies on the skin colour mutation of grapevine have been conducted on the black-skinned Pinot Noir and the white-skinned Pinot Blanc (Yakushiji et al. 2006). Pinot noir, heterozygous for *VvmybA1*, comprises a functional allele, capable of anthocyanin expression, and a non-functional allele, which has lost its capability by the insertion of *Gret1*, whereas Pinot blanc considered to be arisen from the dark-skinned Pinot noir, possesses only a non-functional allele lacking the functional part of the gene. Studies on the identification and isolation of this null-allele in Pinot blanc are in progress (Yakushiji et al. 2006). Moreover, it has been observed that recombination between the LTRs of *Gret1* have lead to solo LTRs or even to a total loss of the retrotransposon in coloured cultivars. This excision event resulted in new dark-skinned varieties originating from white progenitors such as Red Chardonnay, which is derived from Chardonnay (Kobayashi et al. 2004, Yakushiji et al. 2006).

Retrotransposons have repeatedly been used for studying polymorphisms among grapevine cultivars and clones, and have revealed promising results. Relying on the presence of retrotransposon reverse transcriptase sequences, inverse sequence-tagged repeat analyses (ISTR) have been conducted for investigating genetic diversity among closely related *Sangiovese* accessions. ISTR fingerprints provided a high level of polymorphism whereby clonal distinction was successful (Sensi et al. 1996). S-SAP analyses, implementing primers based on the LTRs of *Vine-1*, were successful in distinguishing particular clones such as Traminer clones. But the distinction of *Pinot* clones failed, indicating different clonal variability in different cultivars (Imazio et al. 2002, Labra et al. 2003). Pereira et al. (2005) utilised molecular markers based on LTRs of *Gret1* for REMAP and IRAP analyses. In this study, polymorphism among Portuguese cultivars was revealed leading to a successful identification, while the techniques failed in finding polymorphisms between clones of the same cultivars. Pelsey et al. (2003) have assessed the discriminative power of S-SAPs, relying

on the LTRs of grapevine retrotransposons within 12 *Vitis vinifera* varieties. They confirmed their efficiency in distinguishing each variety from one another (Pelsey et al. 2003).

Transposition seems to play a significant role in the generation of clonal variation. Results of the analysis, applying a S-SAP approach combining *Mse*-primers with universal retrotransposon primers analyzing six Pinot clones in two replications, provided evidence that the similarity among these clones (despite the lower number of genotypes employed) is lower than in comparable AFLP-studies (Wegscheider et al. 2008).

First global and detailed results on the abundance of class II TEs have been presented, based on *in silico* analysis of the publicly available sequences of the *Vitis* genome (Jaillon et al. 2007; Velasco et al. 2007). Over 1160 potentially complete grapevine transposons as well as more than 2000 defective copies were characterized representing approximately 2.0 % of the grapevine genome (Benjak et al. 2008). The same study confirmed that the TE activity highly contributed to the *Vitis* genomic variability. Morover, some TE families have functional copies and are transcriptionally active. Another study of Benjak et al. (in prep), showed various insertion polymorphisms among grapevine cultivars that was caused by MITEs inserting within genes. These insertions give rise to different transcripts that can be found only in some cultivars, suggesting that insertion polymorphisms are very likely linked to phenotypic variation (Lippman et al. 2004). *In silico* analysis furthermore confirmed that expression of TEs in *Vitis* spp. is induced by stress (Benjak et al. 2008).

Clonal variation by chimerism

Chimeric grapevines, in particular periclinal chimeras, have been observed in the past and implemented in clonal selection programs. By convention, periclinal chimera have a two-layered-tunica above a corpus with one or more genetically different apical cell layers. Each of these cell layers remains developmentally independent from the adjacent layers. Because of the stratified meristem morphology, most somatic mutations are not fatal (Hocquiny et al. 2004). This structure is a stabilized chimeric form and can be maintained and amplified through vegetative propagation. Somatic mutation events may be induced within these meristematic layers through either mutated cells deriving from the initial shoot meristem,

or mutated soma cells that may be incorporated into an adventitious meristem, which then develops into a shoot for the mutant phenotype. First descriptions of chimeric grapevine phenotypes were reported in the middle of the 19th century describing red and white colored Pinot clusters occurring on one vine (e.g. Breider 1967). Molecular analysis has added proof through recent genetic studies covering periclinal cytochimeras for Gamay (Thomson and Olmo 1963) to grapevine bud sports resulted in multiple colored grapes (e.g. Walker et al. 2006, Hocquigny 2004). Chimerism has been confirmed by molecular marker studies employing microsatellite markers within the Pinot group (Hocquigny et al.2004), P. meunier (Franks et al. 2002), Chardonnay (Riaz et al.2002) and Greco di Tufo (Crespan et al. 2006) as chimeras through the presence of a third or fourth microsatellite allele. There is evidence that chimeric clones arise from many grapevine cultivars (e.g. Riaz et al. 2002). A grapevine cultivar has rarely been identified to be periclinal chimeric, such as *V. vinifera* cv. Pinot meunier (Skene and Barlass 1983). The P. meunier phenotype has hairy leaves but eventually exhibits mutations (loss of trichomes on leaf surfaces). Studies found P. meunier to be tri-allelic at several loci instead of the usual di-allelic genotypes in grapevines. Pinot meunier shares two alleles with Pinot noir at the locus VVS2 (138:153bp) plus one additional allele (129bp) (Franks et al. 2002). The underlying gene mutation is not yet clearly identified. Stenkamp et al. (2008) studied clonal variation of chimeric Pinot meunier employing 11 Pinot meunier wild type clones of various origins in comparison with mutated genotypes of various ages by AFLP-PCR. Eighteen primer combinations generated a total of 670 reproducible bands of which 161 (24.02 %) were polymorphic. Variation (presented as the percental rate of interclonal polymorphisms, Stenkamp et al. 2008) among all samples of both groups of wildtype and mutated P. meunier clones (mean 1.5 % per sample) showed to be of similar as in other Pinot varieties (Blaich et al. 2007). Interestingly, the variation among clones was higher in the non chimeric mutations (1.3 %) than in the chimeric wild type (0.6 %) confirming the stability of a periclinal chimera (Stenkamp et al. 2008). However this study showed also, that chimeric forces contribute only partly to the overall clonal variation in grapevine. This is in accordance to an extensive genetic variation experiment on five “Pinot” cultivars (P. noir, P. gris, P.blanc, P. meunier, P. moure) performed by Hocquigny et al. (2004) who propose a common, diallelic ancestors (genotype I) for all five Pinot cultivars and show experimental evidence that divergent genotypes have arisen from this

genotype through differential mutation accumulations and cell layer arrangements driven by yet unknown chimeric forces.

29.4 Sex is not lost in grapevine but is rare

Sexual reproduction is employed for breeding purposes and genetic research in grapevine. A short current status quo on grapevine genetics and molecular breeding is introduced. The grape cultivars currently in worldwide cultivation are mostly the centuries-old progeny of vines that mated promiscuously in vineyards and are highly heterozygous (Bowers et al. 1999). Selfing seems to be a rare mechanism of parentage, although grapevine cultivars are hermaphrodites and self-pollinators (di Gaspero et al. 2005). At the molecular level, heterozygosity manifests itself in DNA sequence divergence among the different species and between cultivars and clones of *V. vinifera* as evidenced by results from molecular based on genotyping and on sequencing of allelic variants of genes (e.g. Salamaso et al. 2004, Adam-Blondon et al. 2004, Hoffmann et al. 2006). Molecular maps have been developed (e.g. Doligez et al. 2002, Grando et al. 2003, Fischer et al. 2003, Riaz et al. 2004) covering more than 425 Mbp of the grapevine genome. Grapevine heterozygosity is also expected to be existing in the assortment of genes expressed and in the level at which they are transcribed (Fung et al. 2007). Grape ESTs (expressed sequence tags) libraries have been constructed and assembled in a combined effort to facilitate gene discovery, transcription profiling and SNP marker development (e.g. Moser et al. 2005 for a review), providing insight into organ-specific expression of berry, root, leaf, bud, shoot and inflorescences of several grape cultivars. Several EST-banks are open to the scientific community and are used for both applied and molecular breeding efforts (e.g. da Silva et al. 2005). The nucleotide sequence of *V. vinifera* cv. Pinot noir via a shot gun approach has been recently released by Velasco et al. (2007); it will greatly boost further research on grapevine genomics and allow further insights into effects on both sexual and asexual propagation.

29.5 Advantages of grapevine clones

Apart from the already mentioned advantages of clones in grapevine breeding, what are the advantages of clones for growers? The major benefit to growers is certainly the identical genotype of every plant in a vineyard and consequently, identical behaviour and growth stages. Plants of a clone will have their bud burst at the same time; their shoots will grow at the same speed and direction, which makes canopy management much easier. All plants of a clonal vineyard will require crop protection at the same time and at the same dosage, which increases efficiency, reduces costs, the amount of pesticides used and their impact on the environment. At the end of the growing season, all plants of a clonal vineyard will commence ripening simultaneously and be ready for harvest at the same time. So, all grapes can be harvested at the right time with a maximum in quality. Therefore, the use of clonal material has many economical and ecological advantages.

Are there also disadvantages in the use of clones? Looking at a clonal vineyard from an ecological point of view, it is an extreme form of monoculture. Identical genotypes are growing side by side throughout the field, a pest or disease specialised in this genotype could wipe out the whole planting. But so far, there is no evidence that clonal plantings are more threatened by pests and diseases than other varietal plantings. The obvious reason for this is that a vineyard is largely a monoculture anyhow and the genetic differences between clones regarding resistance to pests and diseases are – apart from bunch rot (see figure 1) - very small. Therefore, clones do not contribute to the monoculture character of a vineyard and do not increase the pest and disease risk. In the case of botrytis, a tolerant clone is far better than a mixed clonal planting.

Do grapevine clones make wines better?

This is a very difficult question to answer, as quality and wine quality in particular cannot be measured. Quality is a subjective term. Consequently, wine drinkers very often completely disagree on the quality of a wine. While we cannot measure quality itself, we can measure quality parameter, e.g. acidity, sugar, alcohol, colour or aromas. That leads us to the question whether clones can influence quality parameters. As we have already seen (Figure

1), different clones can produce wine with differences in acidity or sugar content. So when a wine grower chooses a clone he/she can decide on a particular wine type by selecting a clone with the required quality parameters. If the vineyard is in a humid area, it might be a good idea to choose a clone with loose clusters rather than one with tight compact bunches, being highly susceptible to bunch rot. In these cases, clones are certainly a measure to improve quality. But what about other parameters like flavour and complexity? Do not wines from single clones lack flavour and complexity? Particularly in countries where clones are fairly new, the wines of new clonal plantings are often reported as lower in quality. When a high yielding clone was planted, this is quite obvious. With other clones, the reason for these reports is that new clonal vineyards are compared with old non-clonal vineyards and that the wine quality of a vineyard usually increases with age. The reason for quality differences is the different age of the vineyards and not the use of clones. Consequently, apart from the economic and ecologic advantages mentioned earlier, clones can also be used to increase wine quality. Therefore, clones are increasingly used in viticulture worldwide.

29.6 Conclusion

Continuous asexual propagation is the basis for the clonal selection of superior clones in woody perennial crops such as grapevine. A term, describing clones in woody perennial crops could be “a clone is the assemblage of biotypes deriving from a single zygote through somatic mutations of various kinds and thus expose genetic and phenotypic variation”.

Selection of grapevine clones combines the search for somatically derived genomic variation to produce new clones with novel traits as well as the elimination of less favourable mutations of existing clones. Clonal selection uses both phenotypic and genotypic markers to select for new and to sustain existing grapevine clones. Research on asexually propagated grapevine clones will progress. New innovative techniques will facilitate closer looks into the variation inducing mechanisms of the grape’s genome. Our view will be expanded to the other plant genomes (mitochondrial, chloroplast) with new techniques. To successfully select for superior grapevine clones, long-term field studies and both fruit and wine analyses are required to be combined with molecular marker studies or metabolomic profiling to reach for

the molecular mechanisms involved in creating clonal variation. Though asexual propagation is not applied in viticulture, the breeder's efforts are essential for our understanding of the grapevine genome. Looking at the position and timing of the mutations will help to find ways of manipulating variation inducing events. Furthermore, quantification of such variation will be of great interest.

Acknowledgements

We greatly appreciate the ongoing interest and discussion on grapevine clonality with our colleagues and students from University of Natural Resources and Applied Life Science, Vienna, Austria and Geisenheim Research Centre, Geisenheim, Germany. This work has been supported by a scholarship awarded to A.F. from the ESF- Parthenogenesis Network.

29.7 References

Adam-Blondon AF, Roux C, Claux D, Butterlin G, Merdinoglu D, This P (2004) Mapping 245 SSR markers on the *Vitis vinifera* genome: a tool for grape genetics. Theor Appl Genet 109:1017–1027

Ambrosi H, Becker H (1978) Der deutsche Wein. Gräfe und Unzer Verlag München

Aradhya MK, Dangl GS, Prins BH, Boursiquot J-M, Walker MA, Meredith CP, Simon CJ (2003) Genetic structure and differentiation in cultivated grape, *Vitis vinifera* L. Genet Res 81:179–192

Benjak A, Konradi J, Blaich R, Forneck A (2006) Different DNA extraction methods can cause different AFLP profiles in grapevine (*Vitis vinifera* L.). Vitis 45: 15-21

Benjak A, Forneck A, Casacuberta JM (2008) Genome-Wide Analysis of the "Cut-and-Paste" Transposons of Grapevine. PLoS ONE, 3(9): e3107

Bennetzen JL (2000) Transposable element contributions to plant gene and genome evolution. Plant Mol Biol. 42: 251-269

Billiard R (1913) La vigne dans l'Antiquité. Laffitte, Lyon, France

Blaich R, Konradi J, Rühl R, Forneck A (2007) Assessing genteic variation among Pinot noir (*V. vinifera* L.) clones with AFLP markers. Am. J. Ecol. Vitic. 58:4:526-529

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

Boidron R, Boursiquot JM, Doazan JP, Leclair P, Leguay M, Walter B (1997) Catalogue of selected wine grape varieties and clones cultivated in France. ENTAV Le Grau du Roi, France

Boss PK, Thomas MR (2002) Association of dwarfism and floral induction with a grape "green revolution" mutation. Nature 416: 847-850

Boursiquot JM, Lacombe T, Bowers J, Meredith C (2004) Le Gouais, un cépage clé du partimoine viticola européen. Bull de l'O.I.V. 77: 875-876: 5-19

Bowers JE, Meredith CP (1997) The parentage of a classic wine grape, Cabernet Sauvignon. Nature Genetics 16: 84-87

Bowers J, Boursiquot JM, This P, Chu K, Johansson H, Meredith C (1999) Historical genetics: The parentage of Chardonnay, Gamay, and other wine grapes of northeastern France. Science (Washington D.C.) 285: 5433: 1462-1565

Breider H (1967) Über eine Mutation bei der Sorte Würzburg B 48-21-4. Rebe und Wein 20: 231-232

Caposella A, Silander JA, Jansen RK, Bergen B, Talbot DR (1992) Nuclear ribosomal DNA variation among ramets and genets of white clover. Evolution 46:1240-1247

Credi R, Babini AR (1997) Effects of virus an virus-like infections on growth, yield, and fruit quality of Albana and Trebbiano Romagnolo grapevines. Am. J. Enol. Vitic. 48:7-12

Crespan M (2004) Evidence on the evolution of polymorphism of microsatellite markers in varieties of *Vitis vinifera* L. Theor App Genet 108(2): 231-237

Crespan M, Cabello F, Giannetto S, Ibanez J, Karoglan Kontic J, Maletic E, Pejic I, Rodriguez-Torres I, Antonacci D (2006) Malvasia delle Lipari, Malvasia di Sardegna, Greco di Gerace, Malvasia de Sitges and Malvasia dubrovacka – synomyms of an old and famous grape cultivar. Vitis 45: 69-73

Columella LJM (60 b. Chr.) De Re Rustica, Vol III. (German Translation:Ahrens, K. (1972) Columella über Landwirtschaft. Akademie Verlag, Berlin.)

Di Gaspero G, Cipriani G, Marrazzo MT, Andreetta D, Prado Castro MJ, Peterlunger E, Testolin, R (2005) Isolation of (AC) n-microsatellites in *Vitis vinifera* L. and analysis of genetic background in grapevines under marker assisted selection. Mol Breed 15:11–20

Doligez A, Bouquet A, Danglot Y, Lahogue F, Riaz S, Meredith CP, Edwards KJ, This P (2002) Genetic mapping of grapevine (*Vitis vinifera* L.) applied to the detection of QTLs for seedlessness and berry weight. Theor Appl Genet 105:780–795

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

Finnegan DJ (1989) Eukaryotic transposable elements and genome evolution. Trends Genet 5: 103-107

Fischer BM, Salakhutdinov I, Akkurt M, Eibach R, Edwards KJ, Töpfer R, Zyprian EM (2004) Quantitative trait locus analysis of fungal disease resistance factors on a molecular map of grapevine. Theor Appl Genet 108:501–515

Forneck A (2005) Plant Breeding: Clonality - A concept for stability and variability during vegetative propagation. In Progress in Botany. ULK Esser, W Beyschlag, J Murata (Ed.), pp. 165-183. Springer Berlin Heidelberg, Germany

Franks T, Botta R, Thomas MR (2002) Chimerism in grapevines: implications for cultivar identity, ancestry and genetic improvement. Theoretical and Applied Genetics 104): 192-199

Froelich GA (1900) Zur Hybridisation der Reben und der Auswahl der Zuchtreben. Weinbau und Weinhandel 18: 230-231

Fung RWM, Wenping Q, Yincai S, Schachtman DP, Hppert K, Fekete K, Kovacs LG (2007) Gene expression variation in grapevine species Vitis vinifera L. and Vitis aestivalis Michx. Genetic Resources and Crop Evolution 54: 1541-1553

Gill DE, Chao L, Perkins SL, Wolf JB (1995) Genetic Mosaicism in Plants and clonal Animals. Annual Review of Ecology and Systematics: 26:423-444.
Grando MS, Bellin D, Edwards KJ, Pozzi C, Stefanini M, Velasco R. (2003) Molecular linkage maps of Vitis vinifera L. and Vitis riparia Mchx. Theor Appl Genet, 106, 1213-1224.

Hocquigny S, Pelsy F, Dumas V, Kindt S, Heloir MC, Merdinoglu D (2004) Diversification within grapevine cultivars goes through chimeric states. Genome 47: 579–589

Hoffmann P, Vaclavicek A, Blaich R, Forneck A. (2006) A SNP Marker can differentiate loose clustered Clones of Pinot noir (*V.vinifera* L.). In: Peterlunger et al.: International Conference on Grape Genetics and Breeding, 2-6th July 2006, Udine, Italy

Imazio S, Labra F, Grassi, M. Winfield M, Bardino M, Scienza A (2002) Molecular tools for clone identification: the case of the grapevine cultivar 'Traminer'. Plant Breed. 121: 531-535

Jaillon O, Aury JM, Noel B, Policriti A, Clepet C, Casagrande A, Choisne N, Aubourg S, Vitulo N, Jubin C, Vezzi A, Legeai F, Hugueney P, Dasilva C, Horner D, Mica E, Jublot D, Poulain J, Bruyère C, Billault A, Segurens B, Gouyvenoux M, Ugarte E, Cattonaro F, Anthouard V, Vico V, Del Fabbro C, Alaux M, Di Gaspero G, Dumas V, Felice N, Paillard S, Juman I, Moroldo M, Scalabrin S, Canaguier A, Le Clainche I, Malacrida G, Durand E, Pesole G, Laucou V, Chatelet P, Merdinoglu D, Delledonne M, Pezzotti M, Lecharny A, Scarpelli C, Artiguenave F, Pè ME, Valle G, Morgante M, Caboche M, Adam-Blondon AF, Weissenbach J, Quétier F, Wincker P, (2007) The grapevine genome sequence suggests ancestral hexaploidization in major angiosperm phyla. Nature: 27:449(7161):463-7.

Klekowski EJ, Godfrey PJ (1989) Aging and mutation in plants. Nature: 340:389-391

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

Konradi J, Blaich R, Forneck A (2007) Genetic Variation among clones and sports of Pinot noir (Vitis vinifiera L.) Eur. J. Hortic. Sci 72: 275-279.

Labra M, Imazio F, Grassi F, Rossoni M, Sala F. (2004) Vine-1 retrotransposon-based sequence-specific amplified polymorphism for *Vitis vinifera* L. genotyping. Plant Breed. 123: 180-185

Levadoux L (1956) Les populations sauvages et cultivées de *Vitis vinifera* L. Ann. Amélior. Plant. (Paris) 6: 59–118

Linder B, Konrad H, Rühl EH (1999) Spätburgunder - Neue Klone praxisreif. Das Deutsche Weinmagazin: 8: 28-31

Lippman Z, Gendrel AV, Black M, Vaughn MW, Dedhia N, McCombie WR, Lavine K, Mittal V, May B, Kasschau KD, Carrington JC, Doerge RW, Colot V, Martienssen R (2004) Role of transposable elements in heterochromatin and epigenetic control. Nature 430: 471-476.

Mannini F, Credi R, Argaimante N (1994) Changes in field performance of clones of grapevine cv. Nebbiolo after virus elimination by heat therapy. In: 6th International Symposium on Grape Breeding, Yalta, Crimea, Ukraine, 4-10 September 1994

Mannini F, Gerbi V, Credi R (1998) Heat-treated virus-infected grapevine clones: Agronomical and enological modifications. Acta Horticulturae: 473:155-163

Moser C, Segala C, Fontana P, Salakhudtinov I, Gatto P, Pindo M, Zyprian E, Toepfer R, Grando MS, Velasco R (2005) Comparative analysis of expressed sequence tags from different organs of *Vitis vinifera* L. Funct Integr Genomics 5:208–217

Negrul AM (1946) Origine de la vigne cultivée et sa classification. Tome 1. Ampélographie d'URSS

Oustric J (1994) "Cahiers Techniques". Le Vigneron des Côtes-du-Rhône

Pelsy F, Merdinoglu D (2002) Complete sequence of Tvv1, a family of Ty1 copia-like retrotransposons of *Vitis vinifera* L reconstituted by chromosome walking. Theor. Appl. Genet. 105: 614-621

Pelsy F, Schehrer, L. Merdinoglu, D (2003) Development of grapevine retrotransposon-based molecular markers (S-SAP). Acta Hortic. 603:83-87

Pereira HS, Barão A, Delgado M, Morais-Cecílio L, Viegas W (2005) Genomic Analysis of Grapevine Retrotransposon 1 (Gret1) in Vitis vinifera. Theor. and Appl. Genet. 111:871-878

Regner F, Stadlbauer A, Eisenheld C (1998) Heunisch x Fränkisch, ein wichtiger Genpool europäischer Rebsorten (*Vitis vinifera* L *sativa*). Wein-Wissenschaft 53:114-118

Regner F (1999) Genetische Analyse von Rebsorten. Deutsches Weinbau-Jahrbuch 51: 125-132

Regner F, Stadlbauer A, Eisenheld C, Kaserer H (2000) Genetic relationships among Pinots and related cultivars. Am. J. Enol. Vitic. 51: 7-14

Riaz S, Garrison KE, Dangl GS, Boursiquot JM, Meredith CP (2002) Genetic divergence and chimerism within ancient asexually propagated winegrape cultivars. Journal of the American Society for Horticultural Science 127(4): 508-514

Riaz S, Dangl GS, Edwards KJ, Meredith CP (2004) A microsatellite marker based framework linkage map of *Vitis vinifera* L. Theor Appl Genet 108:864–872

Rühl EH, Konrad H, Schmitt K, Pfeiffer W (2000) Clonal variation in Pinot Noir. 5th International Symposium on Cool Climate Viticulture and Oenology, 16-20 January, 2000, Technical Paper

Sensi E, Vignani, R, Rohde W, Biricolti, S. (1996) Characterization of genetic biodiversity with *Vitis vinifera* L. Sangiovese and Colorino genotypes by AFLP and ISTR DNA marker technology. Vitis 35: 183-188

Skene KGM, Barlass M (1983) Studies on the fragmented shoot apex of grapevine. IV. Separation of phenotypes in a periclinal chimera in vitro. Journal of Experimental Botany 34(147): 1271-1280

Sievers E (1971) Anatomical differences in clones of the White Riesling cv. (Anatomische Unterschiede bei Klonen der Rebsorte Weisser Riesling). Mitt. Klosterneuburg 21:63-64

Silvestroni O, Intrieri C, Di Domizio N (1995) Valutazione della variabilità clonale in cultivar di *Vitis vinifera* mediante metodi fillometrici. Accademia Italiana della Vite e del Vino, Siena 47, 111-125

Stenkamp S, Becker M, Hill B, Blaich R, Forneck A (2008) Clonal variation and stability assay of chimeric Pinot Meunier (Vitis vinifera L.) and descending sports. Euphytica online first.

Thompson MM, Olmo HP (1963) Cytohistological Studies of Cytochimeric and Tetraploid Grapes. American Journal of Botany 50: 901-906

Tuskan, GA, Francis, KA, Russ, SL, , Romme, WH, Turner, MG (1996) RAPD markers reveal diversity within and among clonal and seedling stands of aspen in Yellowstone National Park, USA. Can. J. For. Res. 26:2088-2098

Velasco R, Zharkikh A, Troggio M, Cartwright DA, Cestaro A, et al (2007) A High Quality Draft Consensus Sequence of the Genome of a Heterozygous Grapevine Variety. PLoS One 2(12): e1326.

Verries, C, Bes, C, This, P, Tesniere, 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

Walker AR, Lee E, Robinson SP (2006) Two new grape cultivars, bud sports of Cabernet Sauvignon, bearing pale-coloured berries, are the result of deletion of two regulatory genes of the berry colour locus. Plant, Cell and Environment. 29: 2185-2197

Walter B, Martelli GP (1996) Clonal selection of the vine: Sanitary and pomological selection. Influence of viruses and quality. Part 1: Effects of viruses on the culture of the vine and its products. Bulletin de l'O.I.V. 68: 945-971

Walter B, Martelli GP (1997) Clonal selection of the vine: Sanitary and pomological selection. Influence of viruses and quality. Part 2: Sanitary selection – Pomological selection. Bulletin de l'O.I.V. 70: 5-23

Wegscheider E, Benjak A, Forneck A (2008) Clonal variability in Pinot noir revealed by S-SAP involving universal retrotransposon-based sequences. Am. J. Enol. Vitic. Accepted.

Yakushiji H, Kobayashi S, Goto-Yamamoto N, Tae Jeong S, Sueta T, Mitani N. Azuma A (2006) A skin color mutation of grapevine, from black-skinned Pinot Noir to white-skinned Pinot Blanc, is caused by deletion of the functional VvmybA1 allele. Biosci. Biotechnol. Biochem. 70: 1506-1508

Ye G, Soylemezoglu NG, Weeden F, Lamboy WF, Pool RM, Reisch BI (1998) Analysis of the relationship between grapevine cultivars, sports and clones via DNA fingerprinting. Vitis 37: 33-38

Zohary D, Spiegel-Roy P (1975) Beginnings of fruit growings in the old world. Science 187: 319–327

ARTICLE 2

Different DNA extraction methods can cause different AFLP profiles in grapevine (*Vitis vinifera* L.)

A. Benjak, J. Konradi, R. Blaich and A. Forneck

Inclusion permission granted by publisher.

Vitis **45** (1), 15–21 (2006)

Different DNA extraction methods can cause different AFLP profiles in grapevine (*Vitis vinifera* L.)

A. Benjak[1], J. Konradi[2], R. Blaich[2] and A. Forneck[2]

[1] University of Natural Resources and Applied Life Sciences, Department of Applied Plant Sciences and Biotechnology, Institute of Horticulture, Fruit-Growing and Viticulture, Vienna, Austria

[2] University of Hohenheim, Department of Special Crop Cultivation and Crop Physiology, Section of Viticulture (370a), Stuttgart, Germany

Summary

Amplified fragment length polymorphism (AFLP) is widely used for DNA fingerprinting and it has been broadly applied in population genetics. Since it is based on restriction digestion and PCR-based amplification it can be influenced by different chemical compounds commonly found in the isolated DNA. DNA extraction procedures may alter the AFLP banding profiles through DNA quality. Hence the DNA extraction method is crucial to produce reproducible AFLP-banding profiles.

In this work two sets of AFLP analyses were performed on 62 Pinot noir, 6 Pinot blanc and 4 Pinot gris (*Vitis vinifera* L.) clones, and profiles obtained after three different DNA extraction methods were compared. AFLP profiles were different for the same genotypes due to the DNA extraction method used.

K e y w o r d s: DNA extraction, *Vitis*, AFLP-PCR.

Introduction

Amplified fragment length polymorphism (AFLP) PCR techniques (Vos *et al.* 1995) are widely used for DNA fingerprinting. AFLP markers can be generated from DNA of any origin, therefore they have been used effectively in bacteria, fungi, animals and plants (Mueller and LaReesa Wolfenbarger 1999), including grapevine (*Vitis vinifera* L.) (*e.g.* Cervera *et al.* 1998, 2000, Goto-Yamamoto 2000, Popescu *et al.* 2002, Vignani *et al.* 2002, Fanizza *et al.* 2003, Forneck 2005).

The quality of the extracted DNA and the method of extraction could affect the profiles obtained (Jones *et al.* 1997, Reineke *et al.* 1998, Botteux *et al.* 1999), because several types of contaminants in the DNA can reduce the activity of restriction endonucleases, polymerases and ligases (Shioda and Marakami-Muofushi 1987, Do and Adams 1991). A complete digestion of DNA is crucial for the accuracy of AFLP fingerprinting. It was found that in excess of restriction enzymes as applied in AFLP procedures, partial digestion may result from star activity in enzymes or contamination of DNA with negatively charged polysaccharides and phenols (Do and Adams 1991, Demeke and Adams 1992, Lodhi *et al.* 1994), usually found in DNA extracted from *Vitis vinifera* L. As an example for a polysaccharide heparin, occurring in animals, was found to inhibit *Eco*RI endonuclease cleavage of DNA at certain *Eco*RI sites (Chen *et al.* 1990). Many factors inhibiting the PCR reaction were determined, including detergents, antibiotics, enzymes, polysaccharides, fats, proteins and other organic and inorganic chemical compounds (Rossen *et al.* 1992, Wilson 1997).

The quality of DNA depends on the extraction method used as well as on the additional purification steps. Reineke *et al.* (1998) reported different AFLP profiles obtained from differently purified DNA from *Lymantria dispar* insects. Apart from the initial DNA extraction method, post extraction DNA purification steps may have additional impact on AFLP profiles. Since many innovative DNA extraction kits routinely apply column-based purification steps in the protocol (*e.g.* Green and Thompson 1999) this may be of relevance for further argumentation. As an example Zhang *et al.* (1999) reported variable AFLP fingerprints in *Rosa ssp.* when using DNA isolated with two different methods (CTAB based and Qiagen DNeasy Plant Mini kit).

Several DNA extraction protocols are commonly used for fingerprinting in grapevine, mostly as modifications of the analog method. A similar extraction buffer based on Tris, EDTA and 2-mercaptoethanol (Thomas *et al.* 1993) or with an addition of cetyltrimethylammonium bromide (CTAB) (*e.g.* Doyle and Doyle 1990, Bowers *et al.* 1993, Lodhi *et al.* 1994, Wolf *et al.* 1999, Labra *et al.* 2001) is usually applied. A recent alternative to these methods is the column based Qiagen DNeasy Plant Mini Kit which yields sufficient good quality DNA; it has already been used for grapevine fingerprinting (*e.g.* Pollefeys and Bousquet 2003, Adam-Blondon *et al.* 2004, This *et al.* 2004).

Vitis vinifera ssp. and related species have been the subject of extensive genetic studies due to their worldwide cultivation and importance. Since AFLPs are frequently used to differentiate closely related genotypes, such as vegetatively propagated, identical "clones", where the genetic polymorphism is low, it is important to be aware of possible modifying factors of any AFLP profile. If DNA extraction methods pose such selection pressure on data, this must be pointed out and in consequence corrected by aligning methods. The goal of our work was to compare AFLP results in closely related grapevine genotypes using three different DNA extraction methods and to detect the most reliable method for AFLP fingerprinting. We are reporting

Correspondence to: Dr. A. Forneck, University of Natural Resources and Applied Life Sciences, Department of Applied Plant Sciences and Biotechnology, Institute of Horticulture, Fruit-Growing and Viticulture, 1190 Vienna, Austria. Fax: +43-1-47654-3359. E-mail: astrid.forneck@boku.ac.at

the occurrence of variable AFLP profiles and statistic analyses in grapevine depending on the DNA extraction method used.

Material and Methods

Two individual analyses with different samples and different AFLP protocols were conducted in this work, further referred to AFLP analysis 1 and AFLP analysis 2.

AFLP analysis 1: Plant material and DNA extraction methods: Six clones of Pinot blanc (2-53Gm, 10-13Gm, 2-21Gm (Forschungsanstalt Geisenheim, Germany), D55, D57, and EA98-04 (Weinbauinstitut Freiburg, Germany)) and 4 clones of Pinot gris (D42, D53, FR52-121 (Weinbauinstitut Freiburg, Germany and H-1 (Hauser-Bühler, Vogtsburg-Bickensohl, Germany)) were analyzed in this work. Total DNA was isolated from young leaves (stored at -20 °C) using three different methods.

Method 1 was a modified CTAB method (with 6 % PVP) from Doyle and Doyle (1990). Samples were ground in liquid nitrogen and dispersed in 700 µl of extraction buffer (0.1 M Tris pH 8.0, 1.4 M NaCl, 2 % (w/v) cetyltrimethyl-ammonium bromide (CTAB), 0.2 % (v/v) 2-mercaptoethanol, 20 mM ethylenediaminetetraacetic acid (EDTA) and 6 % (w/v) polyvinylpyrolidone (PVP)) and incubated at 65 °C for 30 min with occasional mixing by gentle tube inversion. Tubes were kept on ice, 700 µl of chloroform-isoamyl alcohol (24:1, v:v) was added and samples were shaken gently for 20 min, then centrifuged at 14, 000 rpm for 8 min, 600 µl of aqueous phase was removed and 15 µl of RNAse (10 $mg \cdot ml^{-1}$) were added for a 30 min incubation-step at room temperature. 1/10 volume of 3M Na-acetate and 2/3 volumes of ice-cooled isopropanol were added and mixed by gentle inversion. Samples were stored at -20 °C for 20 min than centrifuged at 14,000 rpm for 10 min. The pellet was rinsed with 500 µl of 70 % ethanol, dried at 40 °C and resuspended in TE buffer.

Method 2, a modified protocol based on THOMAS *et al.* (1993), described in BÖHM (2000), did not contain CTAB in the extraction buffer. Two sets of ground samples (using liquid N_2) were dispersed in 1.2 ml of the extraction buffer "A" (0.2 M Tris HCl pH 8.0, 0.25 M NaCl, 0.1 % (v/v) 2-mercaptoethanol, 50 mM EDTA and 2.5 % (w/v) PVP), vortexed and centrifuged for 8 min at 14,000 rpm. The liquid phase was poured and the pellet resuspended in 0.8 ml of the extraction buffer "B" (0.2 M Tris HCl, pH 8.0, 0.5 M NaCl, 50 mM EDTA, 2.5 % (w/v) PVP, 3 % (w/v) Sarkosyl and 20 % (v/v) ethanol) and incubated for 30 min on 37 °C with occasional mixing by gentle tube inversion. An equal volume of chloroform-isoamyl alcohol (24:1, v:v) was added, mixed and centrifuged for 5 min at 14,000 rpm. This step was repeated twice, by collecting the aqueous phase (0.6 and 0.45 ml respectively) and adding one volume of chloroform-isoamyl alcohol (24:1, v:v). A total amount of 0.3 ml of the aqueous phase from the same two samples was pooled into one tube and 0.3 ml of isopropanol was added. After 10 min of centrifugation at 14,000 rpm the aqueous phase was poured and the pellet resuspended in 100 µl TE buffer. RNAse was added following 15 min incubation at room temperature. 100 µl of 7.5 M ammonium acetate, pH 8, was added followed by centrifugation 2 min, 10,000 rpm. The aqueous phase (190 µl) was collected in a new tube together with 190 µl of cold absolute ethanol and incubated for 10 min in the refrigerator, followed by centrifugation (10 min, 10,000 rpm), rinsing the pellet with 70 % ethanol, drying the pellet and resuspending it in 60 µl of TE buffer.

In method 3 Qiagen DNeasy Plant Mini kit was used for DNA extraction following the original procedure of the kit, supplemented by the manufacturer (Qiagen, Hilden, Germany).

DNA concentration was estimated by 1.5 % agarose gel electrophoresis using λ DNA (25, 50, and 100 $ng \cdot \mu l^{-1}$).

AFLP protocol: AFLP analysis was performed according to Vos *et al.* (1995) with the modifications described below. Digestion was carried out in a final volume of 25 µl using the y^+/Tango buffer with BSA (Fermentas, St. Leon-Rot, Germany), 45 U of *Eco*RI, and 3.6 U of *Tru*1I restriction enzymes (Fermentas, St. Leon-Rot, Germany) during 1.5 h at 37 °C followed by 2 h at 65 °C and 15 min at 85 °C. Ligation was done adding 5 µl of a mix containing 5 pmol of *Eco*RI adapter, 50 pmol of MseI adapter, 2 mM ATP, 5 U of T4 DNA ligase and ligation buffer (Fermentas, St. Leon-Rot, Germany). The ligation was incubated overnight at room temperature.

The first amplification was performed in a total volume of 20 µl using 3 µl of digested-ligated DNA template, 10 pmol of each primer, 2 mM of each dNTP, 3 mM $MgCl_2$, 0.3 U of *Taq* DNA polymerase recombinant (Invitrogen, Karlsruhe, Germany) and PCR buffer. The PCR amplifications were carried out applying the following PCR-steps: 94 °C·1 min^{-1} + 26 x (94 °C·30 s^{-1}, 56 °C·1 min^{-1}, 72 °C·1 min^{-1}) + 72 °C·6 min^{-1}. The PCR products were diluted 1:20 and 2 µl were added in total volume of 20 µl PCR reaction containing 10 pmol of each primer, 2 mM of each dNTP, 3 mM $MgCl_2$, 0.5 U of *Taq* DNA polymerase recombinant (Invitrogen, Karlsruhe, Germany) and PCR buffer. The PCR program was a touchdown: 94 °C·min^{-1} + 11 x (94 °C·30 s^{-1}, 65 °C·30 s^{-1} (decreasing 0.8 °C every cycle), 72 °C·min^{-1}) + 26 x (94 °C·30 s^{-1}, 56 °C·30 s^{-1}, 72 °C·min^{-1}) + 72 °C·min^{-1}.

Four primer pairs were used in this analysis, chosen after screening among 16 pairs. One primer in a pair was marked with a fluorescent carbocyanine dye Cy™ 5 (MWG-Biotech AG, Ebersberg, Germany). The pairs were as follows: $E10_{Cy5}$-M16, E16-$M17_{Cy5}$, E(+0)-$M8_{Cy5}$ and $M8_{Cy5}$-$M17_{Cy5}$ (Tab. 1).

Table 1

Sequences of the primers used in this work

Primer	Sequence	Type
E(+0)	5'-GACTGCGTACCAATTC-3'	*Eco*RI
E7	5'-GACTGCGTACCAATTCATG-3'	primers
E10	5'-GACTGCGTACCAATTCACA-3'	
E15	5'-GACTGCGTACCAATTCAGG-3'	
E16	5'-GACTGCGTACCAATTCATC-3'	
M8	5'-GATGAGTCCTGAGTAAATG-3'	*Mse*I
M16	5'-GATGAGTCCTGAGTAAATC-3'	primers
M17	5'-GATGAGTCCTGAGTAAAGT-3'	
M19	5'-GATGAGTCCTGAGTAACAG-3'	

The AFLP technique was confirmed for reproducibility by using standard control samples. Electrophoresis was done on 6 % acrylamide-bisacrylamide (19:1), 6.75 M urea and 0.6 x TBE gels running in 0.5 x TBE buffer on an automated analyzer (ALFexpress™ II DNA Analysis System, Amersham Biosciences, Freiburg, Germany). Bands were displayed and analyzed using Allele Locator 1.03 software (Amersham Biosciences, 1998).

The AFLP analysis 2 was done with 62 samples from Pinot noir clones (Tab. 2), using two DNA extraction methods, method 2 and method 3 described above. Digestion and amplification followed the methods described above with the exception that the primers were not fluorescently labeled; they were synthesized by Invitrogen, Karlsruhe, Germany. For the selective amplification three primer pairs were used: E7-M17, E15-M8 and E16-M19 (Tab. 1). The amplification products were separated on a 6 % polyacrylamide gel at 1600 V and silver stained as described in Bassam and Caetano-Annoles (1993).

Statistic analysis: The statistic analysis for both analyses was done using NTSYS-PC software, version 1.8 (Rohlf 1993). Dendrograms were constructed based on Simple Matching genetic distance and UPGMA clustering following the SAHN procedure (Sneath and Sokal 1973).

Results

AFLP analysis 1: All three DNA extraction methods used yielded sufficient DNA (method 1: mean 1.14 mg DNA, $\delta = 0.76$; method 2: mean 3.54 mg DNA, $\delta = 0.87$; method 3: mean 5 mg DNA , $\delta = 0$). The uniformity of DNA extracted was lowest in method 1 ranging from 0.44-2.75 mg DNA.

The number of total markers found (mean 112) and the degree of polymorphism (21.3 % average) was similar for all methods (Tab 3). Method 3 samples had 4.7 % of missing values in contrast to 2.1 % and 2.9 % for method 1 and method 2 respectively.

Each DNA extraction method produced a different AFLP-banding pattern for the very same genotype. This occurred also in the polymorphic 38 markers found indicating that no

Table 2

Samples used in the AFLP analysis 2

Sample number	Clone name (not official)	Clonal material sourced from	Sample number	Clone name (not official)	Location of clonal selection
52	We 815	Staatliche Lehr- und	156	EA 86-10	Weinbauinstitut
54	We 808	Versuchsanstalt	157	EA 86-13	Freiburg, Germany
55	We 813	Weinsberg,	158	EA 88-17	
64	We M 242	Germany	159	EA 88-18	
68	We M 171		160	EA 88-19	
69	We M 1		161	Fr 52/86	
71	23		163	Fr 54-102	
72	Schneider		164	Fr 10	
80	18Gm	Forschungsanstalt	165	Fr 11	
81	20Gm	Geisenheim, Germany	166	EA 86-3	
85	20-18Gm		185	EA 88-20	
86	1-36-4Gm		190	AT 89.01.25	Martin Auer, Hallau,
87	1-1Gm		191	AT 89.04.06	Switzerland
89	1-44Gm		193	AT 89.07.53	
90	1-58Gm		195	AT 89.09.07	
91	1-86Gm		203	A 87.21.07M	
93	2-4Gm		205	A 68.13.49	
94	2-9Gm		206	A 68.13.50	
95	2-10Gm		207	A 68.14.23	
96	2-6Gm		208	MII/FAW	
97	20-13Gm		209	M1/17/FAW	
100	20-20Gm		211	2/10 FAW	
101	20-26Gm		213	A.OBL.79.01.46	
102	20-27Gm				Etablissement
103	4Gm		215	Pinot 115	National Technique
106	1Gm		217	Pinot 777	pour l'Amélioration
140	108-8Gm		219	Pinot 28	de la Viticulture,
141	1-7-2Gm				France
150	Fr 12 L	Weinbauinstitut	230	Frank 105 S	Reinhard Frank,
152	Fr 13 L	Freiburg, Germany	232	Frank 105	Kenzingen,
154	EA 79-82		233	F. Charisma	Germany
155	EA 91-01		234	F. Classic	

Table 3

Number of total markers and percentage of polymorphic markers and missing values in AFLP analyses using three different DNA extraction methods

Trial	Markers and missing values	DNA extraction method Method 1	Method 2	Method 3
AFLP analysis 1	Total markers	114	112	111
	Polymorphic markers	21.9 %	19.6 %	22.5 %
	Missing values	2.1 %	2.9 %	4.7 %
AFLP analysis 2	Total markers	-	125	133
	Polymorphic markers	-	48.8 %	24.1 %
	Missing values	-	9.9 %	1.8 %

polymorphic marker could be found without having an impact of the extraction method. For each extraction method statistical analysis was performed to asses genetic differences displayed in dendrograms. The total genetic variation differed among the three methods (SM coefficients) from 0.89 - 0.98 in method 1, 0.93 - 0.97 in method 2 and 0.90-0.99 in method 3 (Fig. 1). The altered amplification patterns, derived from alternative amplifications of random sequences within a genome, led to substantial differences within dendrograms (Fig. 1). For example clones 2-21Gm, D42 and H-1 are very close in method 3, but differentiated in the other two methods. In method 2 clones 10-13Gm and 2-21Gm are the closest in the dendrogram, but more differentiated in the other two methods, especially in method 1.

AFLP analysis 2: A pretest comparison among the two extraction methods used (with only few samples electrophoresed in the same gel) showed that method 3 had a better display of higher molecular weight bands and had more monomorphic bands (Fig. 2). Although the samples from method 2 for the pretest were stored at -20 °C we considered the differences may have occurred because of different DNA extraction methods applied.

Three primer pairs were used for the AFLP analysis of 62 Pinot noir clones. Statistical results between the extraction methods do not match in terms of number of polymorphic bands and missing values. Method 2 produced less markers (125) than method 3 (133), more polymorphic markers occurred in method 2 (61) than in method 3 (32) and 9.9 % of total bands were interpreted as missing values in method 2 (1.8 % in method 3). The missing values derived mostly from some samples having all faint or missing bands in some primer pairs.

A cluster analysis of the dataset was done. All samples with missing values in one or more primer pairs were excluded from the similarity analysis, decreasing the total number of samples to 47, but increasing the accuracy of the results. Method 3 had less polymorphism and many samples could not be differentiated. Still there were samples differently clustered and some samples could be referred as identical when using one DNA extraction method, and different when using another method (Fig. 3).

Discussion

It is commonly accepted that the AFLP method is reliable for phenetic distance analysis in grapevine (GOTO-YAMAMOTO 2000, FANIZZA *et al.* 2003, FORNECK 2005), for differentiation of varieties (CERVERA *et al.* 1998, 2000, VIGNANI *et al.* 2002, FOSSATI *et al.* 2001), clones (CERVERA *et al.* 2002, IMAZIO *et al.* 2002, POPESCU *et al.* 2002,) and sports (SCOTT *et al.* 2000). This suggests that confrontation of grapevine cultivars using the AFLP method is reliable as long as the DNA quality and purity remain constant.

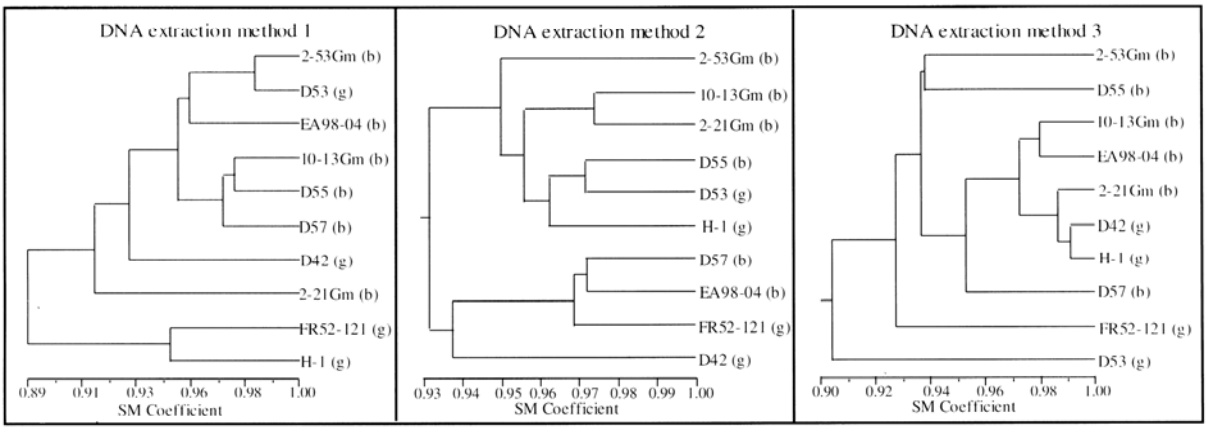

Fig. 1: Dendrograms based on Simple Matching genetic distance and UPGMA clustering for 10 Pinot clones using 3 the DNA extraction methods in AFLP analysis 1. (b) = Pinot blanc, (g) = Pinot gris.

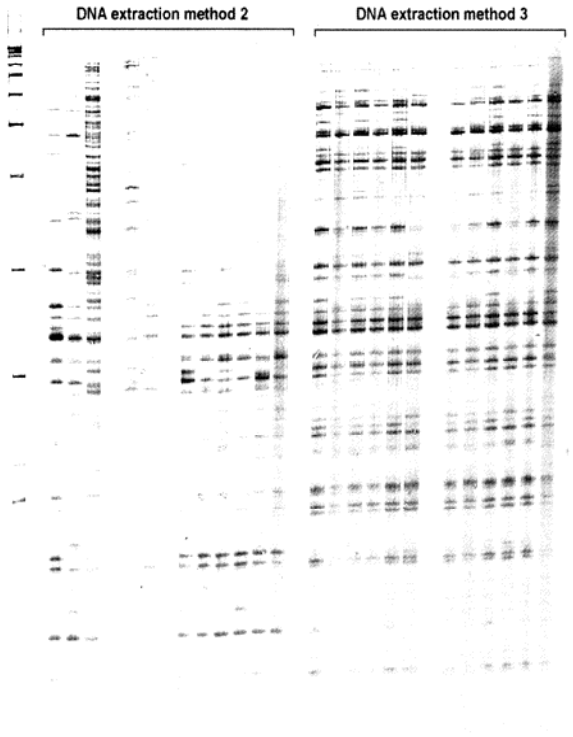

Fig. 2: Example of different AFLP profiles of Pinot noir clones using DNA extraction method 2 and method 3. Bands were displayed using silver staining.

The main prerequisite of restriction polymorphism methods is the complete DNA digestion. The DNA must be as pure as possible for a complete digestion. Since the relative proportions of affecting chemical components vary among cultivars, tissues, and even through seasons a "standard" needs to be found in terms of defining sample tissue and extraction methods. The digestion is usually assessed by gel electrophoresis. However, the critical amount of undigested DNA which could possibly alter the final AFLP results may not be visualized. A suitable method to check for small amounts of undigested DNA is an important issue in AFLP analyses.

In the AFLP analysis 1 only the polymorphic markers were different among extractions used. No monomorphic marker characteristic for one extraction was found (this could not be determined in the AFLP analysis 2 because the samples from two extraction methods were not run together on the gel). The polymorphism detected may be determined by the stable chemical compounds bound on specific sites of DNA making it uncleavable or stopping the PCR amplification at this specific sites. Since these polymorphic bands were reproducibly detected we opt for the occurrence of polymorphism due to DNA structures other than sequence differences or secondary structures such as methylation patterns. We strongly suggest that DNA structures interacting with chemical compounds may alter PCR-based restriction site amplification. Negatively charged polysaccharides and phenols in the DNA may cause partial digestion (Do and Adams 1991, Demeke and Adams 1992, Lodhi *et al.* 1994) or

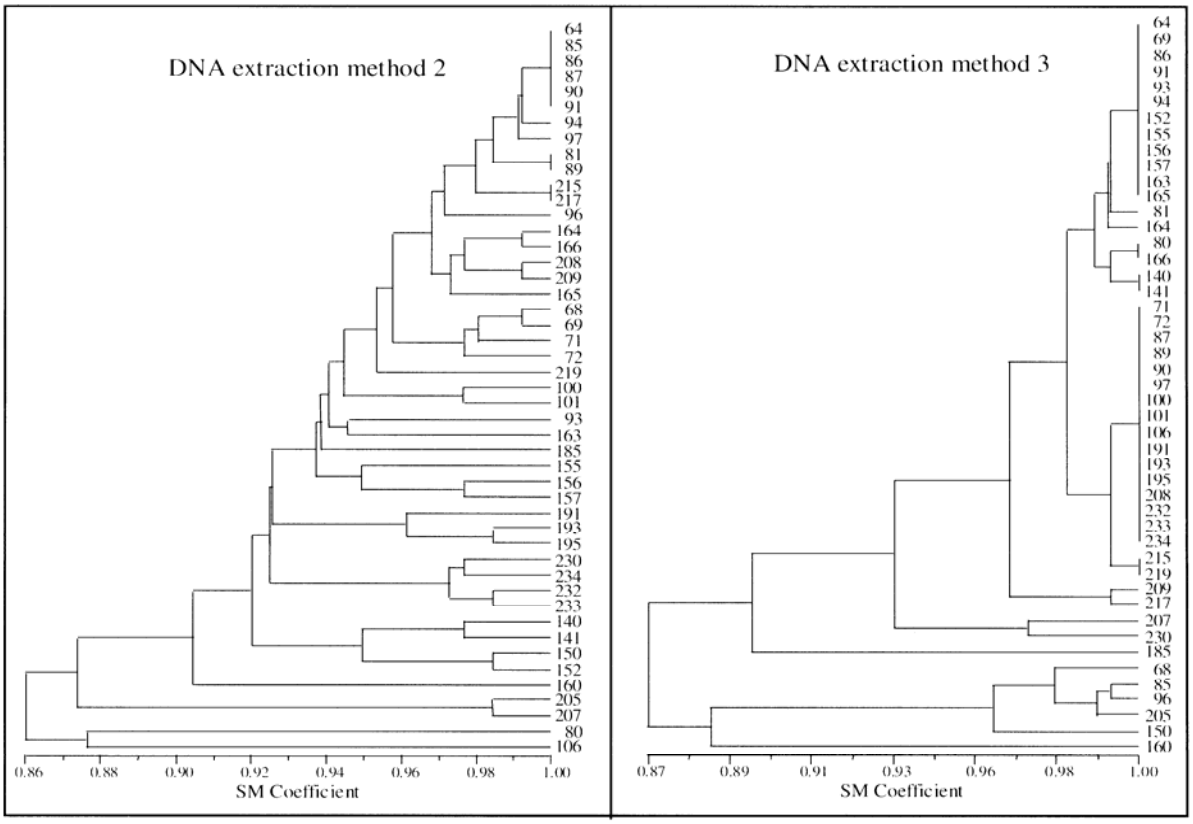

Fig. 3: Dendrograms based on Simple Matching genetic distance and UPGMA clustering in AFLP analysis 2 for 47 Pinot clones using two different DNA extraction methods. The dataset is based on three equal primer pairs.

PCR inhibition (KOONJUL *et al.* 1999). Different extraction methods possibly can differently remove those compounds from the DNA. The high reproducibility of the AFLP and the insensitivity of the procedure to different laboratory conditions have been reported (JONES *et al.* 1997, HANSEN *et al.* 1999, BONIN *et al.* 2004) and we confirm these results for the case of identical DNA extraction methods. Our work shows that the display of some bands in the AFLP profile can be influenced by the DNA extraction method used, therefore combining samples with differently extracted DNA is not recommended. At that point we are unable to specify the reasons of the different results in our AFLP profiles, however, we point out the importance of the DNA extraction method. A top accuracy and fidelity of AFLP profiles is essential especially when fingerprinting closely related genotypes. Due to a higher genetic similarity all factors influencing the accuracy of the band display or inducing intra-genotype polymorphisms might have a bigger impact on final results, thus the whole fingerprinting procedure should be thoroughly standardized.

Different tissue types might have different AFLP profiles (BOITEUX *et al.* 1999, ARANZANA *et al.* 2001, ARNAU *et al.* 2002). This can be due to different degrees of DNA purity obtained from different tissues. Genetic variations due to chimeras might also occur, as was found in the SSR analysis of some grapevine cultivars by Riaz *et al.* (2002). Genetic differences were found among DNA extracted from the same type of tissue on the same plant using AFLP (STENKAMP in prep.) or SSRs (FRANKS *et al.* 2002).

Arnau *et al.* (2002) found irreproducibility in the AFLP due to partial digestion, from tissues sampled in different periods of the growing season and from certain organs.

Another source of genetic variation of a genotype might be transposable elements. They are ubiquistic among all organisms analyzed so far and constitute a large part of plant genomes (KIDWELL and LISCH 1997, BENNETZEN 2000). Here we note that they can be activated in plants by stress (MCCLINTOCK 1984, WESSLER 1996, CAPY *et al.* 2000) changing the original genome sequence. Although the influence of transposable elements was never considered in fingerprinting we think their activity might have repercussions on AFLP results hence the AFLP profile represents equally all parts of the genome analyzed.

Another issue in AFLP analysis is the subjectivity in annotating bands due to disparities in their intensity. BONIN *et al.* (2004) estimate that this error can be 2 % in AFLP analyses. Faint bands were considered as missing values, but the level of intensity between the selective amplification and the background noise is often difficult to standardize. Differences between band display methods might occur. We have compared the two methods used in this work with standard samples and we found no general differences (data not shown). Still it is possible for a band to be faint (annotated as missing value) in one display method and to be more intensive in other methods. This might be especially true for the fluorescent method as it seems to be more sensitive in displaying lower intensity bands. In our results (Tab. 3) a discrepancy in the percentage of missing values occurred between the two experiments. The percentage of polymorphic markers is generally higher in the AFLP experiment 2 than in the experiment 1, especially for the DNA extraction method 2 (48.8 %). The number of samples in the AFLP experiment 2 is bigger (62 *vs.* 10 samples in the AFLP experiment 1) increasing the chances to find polymorphic bands among the samples.

Thus, sampling should be standardized and more samples from the same plant should be verified for differences. Samples should be taken from healthy plants being not under extreme environmental conditions and pathogen free. To reduce statistical errors a larger number of polymorphic bands, excluding the ones containing any missing values, should be used for a better estimation of the genetic distances among genotypes, especially the closely related ones (FANIZZA *et al.* 2003). The fingerprinting procedure should be repeated from the first step. Special care should be taken to decrease human errors, especially the counting and typing of bands, which should be done by two different persons separately (BONIN *et al.* 2004).

Acknowledgements

We thank our colleagues for supplying grapevine samples and sharing their knowledges in numerous discussions with us. We also thank A. SCHREIBER, S. HAVRDA and M. SCHUSTER for technical assistance. The research project at the University of Hohenheim was financially supported by the Forschungsring des Deutschen Weinbaus, Germany.

References

ADAM-BLONDON, A. F.; ROUX, C.; CLAUX, D.; BUTTERLIN, G.; MERDINOGLU, D.; THIS, P.; 2004: Mapping 245 SSR marker on the *Vitis vinifera* genome: A tool for grapevine genetics. Theor. Appl. Genet. **109**, 1017-1027.

ARANZANA, M. J.; DE VICENTE, M. C.; ARÚS, P.; 2001: Comparison of fruit and leaf DNA extracts fro AFLP analysis in Peach (*Prunus persica* (L.) Batsch). Acta Hortic. **546**, 297-300.

ARNAU, G.; LALLEMAND J.; BOURGOIN M.; 2002: Fast and reliable strawberry cultivar identification using inter simple sequence repeat (ISSR) amplification. Euphytica **129**, 69-79.

BASSAM , B. J.; CAETANO-ANOLLES, G; 1993: Silver staining in polyacrylamide gels. Appl. Biochem. Biotechnol. **42**, 181-188.

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

BÖHM, A. 2000: Untersuchungen zur physikalischen Kartierung des Genoms der Weinrebe. Diss. Univ. Hohenheim, Fak. Agrarwissenschaften I.

BOITEUX, L. S.; FONSECA M. E. N.; SIMON P. W.; 1999: Effects of plant tissue and DNA purification method on randomly amplified polymorphic DNA-based genetic fingerprinting analysis in carrot. J. Am. Soc. Hortic. Sci. **124**, 32-38.

BONIN, A.; BELLEMAIN, E.; BRONKEN EIDESEN, P.; POMPANON, F.; BROCHMANN, C.; TABERLET, P.; 2004: How to track and assess genotyping errors in population genetics studies. Mol. Ecol. **13**, 3261-3273.

BOWERS, J. E.; BANDMAN, E. B.; MEREDITH, C. P.; 1993: DNA fingerprint characterization of some wine grape cultivars. Am. J. Enol. Vitic. **44**, 266-274.

CAPY, P.; GASPERI G.; BIEMONT, C.; BAZIN, C.; 2000: Stress and transposable elements: Co-evolution or useful parasites? Heredity **85**, 101-106.

CERVERA, M. T.; CABEZAS J. A.; SANCHA J. C.; MARTINEZ DE TODA F.; MARTINEZ-ZAPATER J. M.; 1998: Application of AFLPs to the characterization of grapevine *Vitis vinifera* L. genetic resources. A case study with accessions from Rioja (Spain). Theor. Appl. Genet. **97**, 51-59.

CERVERA, M. T.; CABEZAS J. A.; SANCHES-ESCRIBANO E.; CENIS J. L.; MARTINEZ-ZAPATER J. M.; 2000: Characterization of genetic variation within table grape varieties (*Vitis vinifera* L.) based on AFLP markers. Vitis **39**, 109-114.

CERVERA, M. T.; CABEZAS J. A.; RODRÍGUEZ-TORRES, I.; CHAVES, J.; CABELLO, F.; MARTINEZ-ZAPATER J. M.; 2002: Varietal diversity within grapevine accessions of cv. Tempranillo. Vitis **41**, 33-36.

CHEN, J.; HERZENBERG L. A.; HERZENBERG L. A.; 1990: Heparin inhibits EcoRI endonuclease cleavage of DNA at certain EcoRI sites. Nucl. Acids Res. **18**, 3255-3260.

DEMEKE, T.; ADAMS R. P.; 1992: The effect of plant polysaccharides and buffer additives of PCR. Biotechniques **12**, 332-334.

DO, N.; ADAMS R. P.; 1991: A simple technique for removing polysaccharides and contaminants from DNA. Biotechniques **10**, 162-166.

DOYLE, J. J.; DOYLE J. L.; 1990: Isolation of plant DNA from fresh tissue. Focus **12**, 13-15.

FANIZZA, G.; CHAABANE, R.; RICCARDI, L.; RESTA, P.; 2003: Analysis of a spontaneous mutant and selected clones of cv. Italia (*Vitis vinifera*) by AFLP markers. Vitis **42**, 27-30.

FORNECK, A.; 2005: Clonality – a concept for stability and variability during vegetative propagation. Progr. Bot. **66**, 164-183.

FOSSATI, T.; LABRA, M.; CASTIGLIONE, S.; FAILLA, O.; SCIENZA, A.; SALA, F.; 2001: The use of AFLP and SSR molecular marker to decipher homonyms and synonyms in grapevine cultivars: the case of the varietal group known as "Schiave". Theor Appl Genet **102**, 200-205.

FRANKS, T.; BOTTA, R.; THOMAS, M. R.; 2002: Chimerism in grapevines: Implications for cultivar identity, ancestry and genetic improvement. Theor. Appl. Genet. **104**, 192-199.

GREEN, M. J.; THOMPSON, D. A.; 1999: Easy and efficient DNA extraction from woody plants for the detection of phytoplasmas by polymerase chain reaction. Plant Dis. **83**, 482-485.

GOTO-YAMAMOTO, N.; 2000: Phenetic clustering of grapes (*Vitis* spp.) by AFLP analysis. Breed. Sci. **50**, 53-57.

HANSEN, M.; KRAFT, T.; CHRISTIANSSON, M.; NILSSON, N. O.; 1999: Evaluation of AFLP in Beta. Theor. Appl. Genet. **98**, 845-852.

IMAZIO, S.; LABRA, M.; GRASSI, F.; WINFIELD, M.; BARDINI, M.; SCIENZA, A.; 2002: Molecular tools for clone identification: the case of the grapevine cultivar "Traminer". Plant Breed. **121**, 531-535.

JONES, C. J.; EDWARDS, K. J.; CASTAGLIONE, S.; WINFIELD, M. O.; SALA, F.; VAN DE WIEL, C.; BREDEMEIJER, G.; VOSMAN, B.; MATTHES, M.; DALY, A.; BRETTSCHNEIDER, R.; BETTINI, P.; BUIATTI, M.; MAESTRI, E.; MALCEVSCHI, A.; MARMIROLI, N.; AERT, R.; VOLCKAERT, G.; RUEDA, J.; LINACERO, R.; VAZQUEZ, A.; KARP, A.; 1997: Reproducibility testing of RAPD, AFLP and SSR markers in plants by a network of European laboratories. Mol. Breed. **3**, 381-390.

KIDWELL, M. G.; LISCH, D.; 1997: Transposable elements as sources of variation in animals and plants. Proc. Natl. Acad. Sci. USA **94**, 7704-7711.

KOONJUL, P. K.; BRANDT, W. F.; FARRANT, J. M.; LINDSEY, G. G.; 1999: Inclusion of polyvinylpyrrolidone in the polymerase chain reaction reverses the inhibitory effects of polyphenolic contamination of RNA. Nucl. Acids Res. **27**, 915-916.

LABRA, M.; CARREÑO-SANCHES, E.; BARDINI, M.; BASSO, B.; SALA, F.; SCIENZA, A.; 2001: Extraction and purification of DNA from grapevine leaves. Vitis **40**, 101-102.

LODHI, M. A.; YE, G. N.; WEEDEN, N. F.; REISCH, B. I.; 1994: A simple and efficient method for DNA extraction from grapevine cultivars, *Vitis* species and *Ampelopsis*. Plant Mol. Biol. Rep. **12**, 6-13.

MCCLINTOCK, B.; 1984: The significance of responses of the genome to challenge. Science **226**, 792-801.

MUELLER, U. G.; WOLFENBARGER, L. L.; 1999: AFLP genotyping and fingerprinting. Trends Ecol. Evol. **14**, 389-394.

POLLEFEYS, P.; BOUSQUET, J.; 2003: Molecular genetic diversity of the French-American grapevine hybrids cultivated in North America. Genome **46**, 1037–1048.

POPESCU, C. F.; FALK, A.; GLIMELIUS, K.; 2002: Application of AFLPs to characterize somaclonal variation in anther-derived grapevines. Vitis **41**, 177-182.

REINEKE, A.; KARLOVSKY, P.; ZEBITZ, C. P. W.; 1998: Preparation and purification of DNA from insects for AFLP analysis. Insect Mol. Biol. **7**, 95-99.

RIAZ, S.; GARRISON, K. E.; DANGL, G. S.; BOURSIQUOT, J. M.; MEREDITH, C. P.; 2002: Genetic divergence and chimerism within ancient asexually propagated wine grape cultivars. J. Am. Soc. Hortic. Sci. **127**, 508-514.

ROSSEN, L.; NORSKOV, P.; HOLMSTROM, K.; RASMUSSEN, O. F.; 1992: Inhibition of PCR by components of food samples, microbial diagnostic assays and DNA-extraction solutions. Int. J. Food. Microbiol. **17**, 37-45.

ROHLF, F. J.; 1993: NTSYS-PC Numerical Taxonomy and Multivariate Analysis System. Version 1.8. Exeter Publications Setauket, New York.

SCOTT, K. D.; ABBLET, E. M.; LEE, L. S.; HENRY, R. J.; 2000: AFLP markers distinguishing an early mutant of Flame Seedless grape. Euphytica **113**, 245-249.

SHIODA, M.; MARAKAMI-MUOFUSHI, K.; 1987: Selective inhibition of DNA polymerase by a polysaccharide purified from slime of *Physarum polycephalum*. Biochem. Biophis. Res. Commun. **146**, 61-66.

SNEATH, P. H. A.; SOKAL, R. R.; 1973: Numerical Taxonomy. W. H. Freeman and Company, San Francisco, CA.

THIS, P.; JUNG, A.; BOCCACCI, P.; BORREGO, J.; BOTTA, R.; CONSTANTINI, L.; CRESPAN, M.; DANGL, G. S.; EISENHELD, C.; FERREIRA-MONTEIRO, F.; GRANDO, S,; IBAÑEZ, J.; LACOMBE, T.; LAUCOU, V.; MAGALHÃES, R.; MEREDITH, C. P.; MILANI, N.; PETERLUNGER, E.; REGNER, F.; ZULINI, L.; MAUL, E.; 2004: Development of a standard set of microsatellite reference alleles for identification of grape cultivars. Theor. Appl. Genet. **109**, 1448-1458.

THOMAS M. R.; MATSUMOTO, S.; CAIN, P.; SCOTT, N. S.; 1993: Repetitive DNA of grapevine: classes present and sequences suitable for cultivar identification. Theor. Appl. Gennet. **86**, 173-180.

VIGNANI, R.; SCALL, M.; MASI, E.; CRESTI, M.; 2002: Genomic variability in *Vitis vinifera* L. "Sangiovese" assessed by microsatellite and non-radioactive AFLP test. Elecronic J. Biotech. **5**, 1-11.

VOS, P.; HOGERS, R.; BLEEKER, M.; REIJANS, M.; VAN DE LEE, T.; HORNES, M.; FRIJTERS, A.; POT, J.; PELEMAN, J.; KUIPER, M.; ZABEAU, M.; 1995: AFLP: A new technique for DNA fingerprinting. Nucl. Acids Res. **23**, 4407-4414

WESSLER, S. R.; 1996: Plant retrotransposons: Turned on by stress. Curr. Biol. **6**, 959-961.

WILSON, I. G.; 1997: Inhibition and facilitation of nucleic acid amplification. Appl. Environ. Microbiol. **63**, 3741-3751.

WOLF, T.; EIMERT, K.; RIES, R.; 1999: Reliable identification of grapevine rootstock varieties using RAPD PCR on woody samples. Austr. J. Grape Wine Res. **5**, 34-38.

ZHANG, D.; GERMAIN, E.; CAO, M.; GRANDELIN, M. H.; 1999: Polymorphism and Reproducibility of AFLPs in rose. Plant and Animal Genome. VIIth Conference, San Diego, USA.

Received September 8, 2005

ARTICLE 3

Clonal variation in Pinot noir revealed by S-SAP involving universal retrotransposon-based sequences

Eva Wegscheider, Andrej Benjak, Astrid Forneck

Manuscript Number: AJEV-D-08-00043R1

Article Type: Research Note

Inclusion permission granted by publisher.

AUTHOR CONTRIBUTIONS:

Conceived and designed the experiments: AB, AF

Performed the experiments: EW

Contributed reagents, material analysis tools: AF

Wrote paper: EW, AB, AF

Clonal variation in Pinot noir revealed by S-SAP involving universal retrotransposon-based sequences

M. sc. Eva Wegscheider, Dipl. ing. Andrej Benjak, Dr. Astrid Forneck

University of Natural Resources and Applied Biotechnology, Institute of Horticulture and Viticulture, Peter Jordan Str. 82, A-1190 Vienna, Austria

Corresponding author: Astrid Forneck email: astrid.forneck@boku.ac.at, University of Natural Resources and Applied Biotechnology, Institute of Horticulture and Viticulture, Peter Jordan Str. 82, A-1190 Vienna, Austria

Key words: Retrotransposon, *Vitis*, clonal variation, S-SAP

Abstract

We present a modified the S-SAP (sequence-specific amplified polymorphism) method using universal primers for retrotransposons for the study on *V. vinifera* cv. Pinot noir. Five Pinot noir clones (20Gm, 1-44Gm, 18Gm, 20-13Gm, 1-84Gm) were analyzed by 30 S-SAP primer combinations employing five MseI- and six universal retrotransposon primers. Altogether 670 markers were generated revealing 4.8 % clonal variation and four out of five Pinot noir clones could be differentiated. This S-SAP method provides an efficient tool to randomly screen for polymorphisms produced through retrotransposition processes in the *Vitis* genome.

Introduction

Because of long term vegetative propagation, a grapevine variety can be composed of a range of clones differing in minor genetic and phenotypic characteristics. One explanation for variation among clones is the occurrence of spontaneous mutations (Forneck 2005). One source of mutations are transposable elements (TEs), which possess the capability to change their genomic location. Thereby they can alter gene structure and rearrange whole genomes

causing major mutational changes (Bennetzen 2000, Kidwell and Lisch 1997). Class I elements, or retrotransposons, transpose via an RNA intermediate, which is reverse-transcribed into cDNA prior insertion into a new target location.

LTR-retrotransposons are composed of LTRs at both ends flanking the internal coding region containing two major genes, *gag* and *pol*. Reverse transcription of the RNA intermediate of a retrotransposon starts at the 5' end of the internal domain, referred as the primer binding site or PBS (Havecker et al 2004), which is a potential target sequence for investigation and detection of retrotransposons by PCR amplifications. LTR-retrotransposons increase their copy number when transposing and contribute to genome size (Kidwell 2002) The copy number, the abundance and the insertion sites of retrotransposons within most investigated plant genomes are considered to be a promising basis for the development of genetic marker systems (Kumar and Hirochika 2001). Several retrotransposon-based marker technologies have been developed to detect a higher degree of polymorphisms at the DNA level. Waugh et al. (1997) established a fingerprinting technique, known as sequence-specific amplified polymorphism (S-SAP). By using sequence-specific retrotransposon primers in combination with AFLP adapter primers a high level of polymorphism is revealed. Kalendar et al. (1999) introduced two retrotransposon-based fingerprinting techniques, Inter-Retrotransposon Amplified Polymorphism (IRAP) and Retrotransposon-Microsatellite Amplified Polymorphism (REMAP), both relying on the positions of LTRs of retrotransposons in the genome. Outward-facing primers, binding to a LTR, are used for the IRAP method, and anneal between two LTRs while primers for REMAP anneal between LTRs and simple sequence repeats. Flavell et al. (1998) introduced a co-dominant marker system based on insertional activities of retrotransposons, known as retrotransposon-based insertion polymorphism (RBIP). The inverse sequence-tagged repeat analysis (ISTR), first implemented by Rohde (1996), relies on the presence of reverse transcriptase sequences of retrotransposons, and has already been used as molecular tool for investigating genetic diversity among closely related grapevine clones (Sensi et al. 1996).

Investigations on clonal variation within grapevine cultivars have shown that the degree of detected genetic divergence usually depends on the applied marker system and on the scope and type of plant samples (Forneck 2005). The retrotransposon-based marker systems S-SAP or ISTR have shown higher levels of polymorphism (Labra et al. 2004, Sensi et al. 1996) than the standard AFLP. Clonal variation studies conducted on *V. vinifera* clones

indicated different levels of clonal variation in different cultivars (e.g. Blaich et al. 2007, Imazio et al. 2002, Regner et al. 2000, Sensi et al. 1996). The main limitation of transposon-based approaches is the need for adequate sequence information for specific primer design. The following study presents a modified transposon display approach based on the original S-SAP procedure (Waugh et al. 1997). The modification employs alternative primers which are universal for several types of plant retrotransposons. Thus, specific sequence information is not required. By applying universal primers we expect to target multiple retrotransposon sequences in the grapevine genome. It is entirely possible to detect an even wider spectrum of mobile elements including all retro-element related sequences which have affected the genome during its evolution. These elements might have survived within intergenic areas or even within coding and regulatory regions such as promoters (Kidwell and Lisch 1997). In fact, many promoters of plant gene sequences contain fragments of transposable elements, indicating a possible contribution to their origin (Wessler et al. 1995, Bennetzen 2000). By targeting a wide range of unknown and also truncated retrotransposon sequences the chance of detecting elements which may have caused changes of evolutionary significance for the grapevine genome, increases. These retrotransposon induced mutations may be also responsible for clonal variation among Pinot noir, resulting in genetic and even phenotypic differences.

In this study, we test the efficiency of a modified S-SAP approach for assaying polymorphism based on retrotransposon sequences. At the same time we assessed the utility of the method for differentiating among Pinot noir clones.

Material and Methods

Plant Material

Six *Vitis vinifera* samples were used for analysis: five clones of the cv. Pinot noir (20Gm, 1-44Gm, 18Gm, 20-13Gm, 1-84Gm) from the Grape Research Station, Germany, and clone ST49 of cv. Domina from the Nursery Steinmann, Sommerhausen/Main, Germany. The Domina clone 'ST49' is the progeny of Pinot noir and Portugieser, and was used for comparison. To confirm the technical reproducibility of the S-SAP method replicates of all six samples were included from the beginning, resulting in a set of 12 samples. Genomic

DNA was extracted from fresh leaves using E.Z.N.A. SP Plant DNA Miniprep Kit according to the manufactorer's instructions (Omega Bio-tek, Doraville, USA).

Transposon Display

Extracted DNA (13.5 µl) was restricted with *Mse*I (Fermentas GmbH, St. Leon-Rot, Germany) in a total volume of 25 µl. The digestion was conducted for 2 h at 65 °C. Restricted DNA was further purified using Perfectprep® Gel Cleanup Kit (Eppendorf, Hamburg, Germany). After purification template DNA (25 µl) was prepared by adding 5 µl of a ligation mix (50 pmol *Mse*I adapter, 100 mM ATP, 10x T4 Ligase buffer and 1 U T4 Ligase (Fermentas GmbH)), and was incubated over night at room temperature (20 °C). T4 Ligase was inactivated by heating up to 65 °C for 10 min, and samples were stored at 4 °C. In the preamplification step the primer M(0) (Table 1), homologous to the adapter sequence, was combined with one of 6 labeled (IRD700 and IRD800) universal retrotransposon primers: F0100, F0103, F0104, F0105, F0113, F0117 (Table 1). The universal transposon primers were obtained from Ruslan Kalendar (MTT/BI Plant Genomics Laboratory, University of Helsinki). These primers were designed from consensus sequences of PBSs of different retrotransposon families deriving from different plant species (Kalendar et al., submitted).

The PCR reaction mixture contained 2.25 µl template DNA, 1.5 µM M(0), 1.5 µM transposon primer, 1x PCR buffer, 3 mM $MgCl_2$, 0.2 mM dNTPs and 1 U Taq DNA polymerase recombinant (Invitrogen Ltd., Paisley, UK) in a final volume of 15 µl. The unselective PCR was conducted using the following program: 94 °C · 60 s^{-1} + 26 x (94 °C· 30 s^{-1}, 56 °C · 60 s^{-1}, 72 °C · 60 s^{-1}) + 72 °C · 6 min^{-1}. The preamplified DNA was diluted (1:10) and stored at 4 °C. The selective amplification was carried out in a total volume of 10 µl containing 1 µl of preamplified DNA, 0.5 µM selective *Mse*I primer (M22, M23, M24, M25, M27) (Table 1), 0.5 µM transposon primer, 1x PCR buffer, 2.5 mM $MgCl_2$, 0.2 mM dNTPs and 0.75 U Taq DNA polymerase recombinant (Invitrogen) using the following cycle profile: 94 °C · 60 s^{-1} + 12 x (94 °C· 30 s^{-1}, 65 °C · 30 s^{-1}, 72 °C · 60 s^{-1}) [the annealing temperature was reduced by 0.7 °C in each of the 12 cycles] + 26 x (94°C· 30 s^{-1}, 56°C · 30 s^{-1}, 72°C · 60 s^{-1}) + 72°C · 6 min^{-1}.

Bands were separated in a 6 % polyacrylamide gel, and visualized by the automated LI-COR NEN 4300 DNA analyzer (Licor Biosciences GmbH, Bad Homburg, Germany).

Sequencing

For further sequence analysis four polymorphic bands (Table 3) were cut out from the gels using the Odyssey Infrared Imaging system (Licor Biosciences GmbH), and stored in 20µl 1 x TE buffer at 4 °C. By three repeated steps of freezing (20 min) and thawing, the DNA was eluted from the polyacrylamide gel into the buffer. 1 µl of each sample was used for a PCR, using M(0) and the corresponding unlabeled transposon primers, to amplify the extracted band. The amplification was conducted using the selective PCR program described above in a total volume of 50 µl. 40 µl of the PCR products were resolved on a 1,5 % agarose (1 x TAE) gel and stained with ROTI-methylene blue staining concentrate (ROTH, Karlsruhe, Germany). The bands, including one repetition for each band, were cut out from the gel, purified with Perfectprep® Gel Cleanup Kit (Eppendorf) and sequenced.

Statistical analysis

The bands were manually scored as present (+) or absent (-). Only reproducible as well as clearly visible bands were recorded. Similarity data matrices were calculated using SM coefficient in the NT-SYS PC program, Version 2.01 (Rohlf 1998). A dendrogram was constructed by SAHN using UPGMA method. DNA sequences were aligned in BioEdit software (Hall 1999), and compared against the nucleotide and protein databases at NCBI using BLAST (Altschul et al. 1990) as well as against the Repbase Update database of transposable elements using the Repeat Masking tool (Jurka et al. 2005).

Results and Discussion

Retrotransposons are characterized by widespread dispersion and various copy numbers within plant genomes. Small genomes, such as *Arabidopsis thaliana* (125 Mb), comprise about 4-8%, while large genomes, such as *Hordeum vulgare* (5000 Mb), host a proportion of approximately 50-80 % retrotransposons (Kumar and Bennetzen 1999). The abundance of retrotransposons in *Vitis vinifera* L. is so far estimated to be from 17 % (Jaillon et al 2007) to 24 % (Velasco et al. 2007) of the genome but detailed annotation of the elements is still not done. The high copy number of retrotransposons in grapevine supports their potential contribution to still undiscovered mutational events, which may have lead to polymorphism among closely related grapevine accessions. We implemented universal retrotransposon

primers in a S-SAP method and revealed 4.8 % overall polymorphism among five Pinot clones tested.

Modified S-SAP procedure for detecting clonal variation

The S-SAP method generated variable banding patterns among the samples (Figure 1). The universal primers were combined with five selective *Mse*I primers (M22, M23, M24, M25, M27) which generated a set of 30 primer combinations. The amplified fragments ranged in size between 50 bp and 350 bp. 98 % of bands were reproducible while the non-reproducible bands were omitted from calculations to ensure consistency.

A total of 670 bands were generated by 30 primer combinations revealing 8.8 % polymorphism (59 polymorphic markers in a total of 670 markers) among all samples. There was 4.8 % polymorphism (32 polymorphic bands) among the Pinot noir clones studied (Table 4) which is higher than if studied with random AFLP-markers utilizing *Mse*I and *Eco*RI primers (Blaich et al. 2007). Based on the presence or absence of amplified fragments, a genetic similarity matrix was calculated using the SM coefficient (Table 2) and an UPGMA based dendrogram was created (Figure 2). The highest rate of polymorphism was obtained by the primer combination F0100 with all five *Mse*I primers (20 out of 174 markers). Generally, the primers F0100, F0103 and F0104 generated significantly more polymorphic markers than F0105, F0113 and F0117 (Table 4). Differences in efficiency may occur since each primer was designed from a different group of plant retrotransposons.

The S-SAP method used here is robust in term of reproducibility and straightforward in the detection of polymorphism. Its specificity is to target a fraction of the genome related to retrotransposons. In a standard AFLP, a reduction in the number of restriction sites to be visualized in a gel (as bands) is needed in order to obtain a banding pattern instead of a smear. This is achieved by using selective primers which result in the visualization of a limited number of restriction sites randomly distributed in the genome. In our method, the retrotransposon-specific primers limit the number of restriction sites to be visualized to only those related to retrotransposons. As retrotransposons are more or less evenly distributed in the non-genic parts of the genome (Brandes et al. 1997), with the S-SAP method used here we also covered the whole genome but less “randomly” compared to standard AFLP. It is possible that transposon related sequences are more prone to mutations compared to

other regions in the genome in general. Visualizing the most variable parts of the genome is the key to detect differences in genetically almost identical grapevine clones.

Elucidating clonal variation among Pinot clones

Clones 20Gm and 20-13Gm were genetically indistinguishable, while the others could be clearly distinguished. Clone 20Gm is a tight clustered Pinot clone from an early selection; whereas Clone 20-13Gm has been sub-cloned after selection treatments and resulted in a small berried phenotype. None of the 670 S-SAP markers correlated with the loose or tight clustered phenotype among the limited set of samples tested. However the clones 1-84Gm and 1-44Gm (newer loose clustered Pinot noir clones) and 18Gm (tight clustered) could be genetically distinguished. Although the overall similarity among the clones studied was generally high (97.5 %), as expected with closely related accessions.

To confirm that the amplified DNA fragments derived from retrotransposon templates, four polymorphic bands, including replicates, were chosen at random and sequenced. Sequences from replicates for each band were identical. The length of the sequenced fragments ranged in size between 160 and 245 bp (Table 3), which is too short for their proper annotation. Therefore we blasted these sequences against the whole genome shotgun reads of the recently sequenced *Vitis vinifera* genome (Jaillon et al. 2007, Velasco et al. 2007) available at NCBI. The results are given in Figure 3. Three sequences out of four derived from mismatching binding of the PBS primers which is in our case an acceptable feature because the primers were designed from PBSs of retrotransposons from other plant species and differences in the PBS sequences are to be expected. All sequences had 98-100 % identity to the *Vitis* sequences available at NCBI. Only one sequence (sequence 2) was not a part of a repetitive sequence and is very close (279 bp downstream) to a putative gene encoding for a putative hydrolase protein. Sequence 1 and 2 belong to unknown repetitive sequences (with low similarities to various class I elements for other plants stored in the Repbase), while sequence 4 is the only one clearly belonging to a LTR retrotransposon. It matches exactly to the 3' of the LTRs which start with the TG and finish with the CA motif followed by 3 nucleotides before the PBS region, which are usual characteristics of LTRs (Suoniemi et al. 1997). The element is flanked by 6 bp target site duplications (TSDs) and from the order of the conserved domains found in the putative *pol* gene it is likely a *copia*-like element. Analysis of the sequence of this element by mean of blastn and blastx shows that the

element is present in several copies, some of which are partially deleted. None of the copies have an intact ORF suggesting that they belong to a family of older and not anymore functional elements.

Two kinds of mutation events could cause a polymorphism in a transposon-based S-SAP: insertion polymorphism caused by the targeted transposon or a mutation in the restriction site specific for the restriction enzyme used. All four polymorphic bands which were sequenced in this study appear to derive from restriction site polymorphism rather than from insertion polymorphism of the repetitive elements which suggests that the retrotransposon activity is difficult to detect using this approach.

Acknowledgements

This research project (04HS021/ 04HS022/ 04HS023) has been financially supported by the Federal Ministry of Food, Agriculture and Consumer Protection of Germany through the Agency for Agriculture and Food (BLE). This work contains part of the master research of E.W. and is part of the PhD research of A.B.

We wish to thank Dr. R. Kalendar for fruitful discussions and for providing unpublished primer sequences.

References

Altschul, S.F., W. Gish, W. Miller, E. W. Myers, and D. J. Lipman 1990. Basic local alignment search tool. J. Mol. Biol. 215, 403-410.

Bennetzen, J. L. and E. A. Kellogg. 1997. Do plants have a one-way ticket to genomic obesity? Plant Cell 9: 1509-1514.

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

Blaich. R., J. Konradi, E. Rühl and A. Forneck. 2007. Assessing Genetic Variation among Pinot noir (*Vitis vinifera* L.) clones with AFLP Markers. Am. J. Enol. Vitic. 58: 526-529.

Brandes, A., J. S. Heslop-Harrison, A. Kamm, S. Kubis, R. L. Doudrick and T. Schmidt. 1997. Comparative analysis of the chromosomal and genomic organization of Ty1-copia-like retrotransposons in pteridophytes, gymnosperms and angiosperms. Plant Mol. Biol. 33: 11-21.

Flavell, A., M. R. Knox, S. R. Pearce and T. H. N. Ellis. 1998. Retrotransposon-based insertion polymorphisms (RBIP) for high throughput marker analysis. The Plant Journal. 16: 643-650.

Forneck, A. 2005. Plant Breeding: Clonality - A concept for stability and variability during vegetative propagation. *In* Progress in Botany. U. L. K. Esser, W. Beyschlag, J. Murata (Ed.), pp. 165-183. Springer Berlin Heidelberg, Germany.

Havecker, E. R., X. Gao and D. F. Voytas (2004) The diversity of LTR retrotransposons. Genome Biol. 5(6): 225.

Imazio, S., M. Labra, F. Grassi, M. Winfield, M. Bardini and A. Scienza. 2002. Molecular tools for clone identification: the case of the grapevine cultivar 'Traminer'. Plant Breeding 121: 531-535.

Jaillon, O., J. M. Aury, B. Noel, A. Policriti, C. Clepet, A. Casagrande, N. Choisne, S. Aubourg, N. Vitulo, C. Jubin, A. Vezzi, F. Legeai, P. Hugueney, C. Dasilva, D. Horner, E. Mica, D. Jublot, J. Poulain, C. Bruyere, A. Billault, B. Segurens, M. Gouyvenoux, E. Ugarte, F. Cattonaro, V. Anthouard, V. Vico, C. Del Fabbro, M. Alaux, G. Di Gaspero, V. Dumas, N. Felice, S. Paillard, I. Jurman, M. Moroldo, S. Scalabrin, A. Canaguier, I. Le Clainche, G. Malacrida, E. Durand, G. Pesole, V. Laucou, P. Chatelet, D. Merdinoglu, M. Delledonne, M. Pezzotti, A. Lecharny, C. Scarpelli, F. Artiguenave, M. E. Pe, G. Valle, M. Morgante, M. Caboche, A. F. Adam-Blondon, J. Weissenbach, F. Quetier and P. Wincker. 2007. The grapevine genome sequence suggests ancestral hexaploidization in major angiosperm phyla, Nature 449 (7161): 463-467.

Kalendar, R., T. Grob, M. Regina, A. Suoniemi and A. Schulman. 1999. IRAP and REMAP: two new retrotransposon-based DNA fingerprinting techniques. Theor. Appl. Genet. 98: 704-711.

Kapitonov, V. V. and J. Jurka. 1999. Molecular paleontology of transposable elements from Arabidopsis thaliana. Genetica 107: 27-37.

Kidwell, M. G. and D. Lisch. 1997. Transposable elements as sources of variation in animals and plants. Proc. Natl. Acad. Sci. U S A 94: 7704-7711.

Kidwell, M. G. 2002. Transposable elements and the evolution of genome size in eukaryotes. Genetica 115: 49-63.

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

Kumar, A. and J. Bennetzen. 1999. Plant retrotransposons. Annual Review of Genetics 33: 479-532.

Kumar, A. and H. Hirochika. 2001. Applications of retrotransposons as genetic tools in plant biology. Trends in Plant Science 6: 127-134.

Labra, M., F. Imazio, F. Grassi, M. Rossoni and F. Sala. 2004. *Vine-1* retrotransposon-based sequence-specific amplified polymorphism for *Vitis vinifera* L. genotyping. Plant Breeding 123: 180-185.

Mao, L., T. C. Wood, Y. Yu, M. A. Budiman, J. Tomkins, S. Woo, M. Sasinowski, G. Presting, D. Frisch, S. Goff, R. A. Dean and R. A. Wing. 2000. Rice transposable elements: a survey of 73,000 sequence-tagged-connectors. Genome Res. 10: 982-990.

Pelsy, F. and D. Merdinoglu. 2002. Complete sequence of Tvv1, a family of Ty1 copia-like retrotransposons of Vitis vinifera L., reconstituted by chromosome walking. Theor. Appl. Genet. 105: 614-621.

Regner, F., A. Stadlbauer, C. Eisenheld and H. Kaserer. 2000. Genetic relationships among Pinots and related cultivars. Am. J. Enol. Vitic. 51: 7-14.

Rohde, J. 1996. A novel PCR-based DNA marker technology called inverse sequence tagged repeat (ISTR) analysis. Genet. Breed. 50: 249-261.

Rohlf, F. J. 1998. NTSYSpc. Numerical Taxonomy and Multivariate Analysis System, Version 2.01.

Sensi, E., R. Vignani, W. Rohde and S. Biricolti. 1996. Characterization of genetic biodiversity with *Vitis vinifera* L. Sangiovese and Colorino genotypes by AFLP and ISTR DNA marker technology. Vitis 35: 183-188.

Suoniemi, A., D. Schmidt and A. H. Schulmann. 1997. BARE-1insertion site preferences and evolutionary conservation of RNA and cDNA processing sites. Genetica 100: 219-230.

Velasco, R., A. Zharkikh, M. Troggio, D. A. Cartwright, A. Cestaro, D. Pruss, M. Pindo, L. M. Fitzgerald, S. Vezzulli, J. Reid, G. Malacarne, D. Iliev, G. Coppola, B. Wardell, D. Micheletti, T. Macalma, M. Facci, J. T. Mitchell, M. Perazzolli, G. Eldredge, P. Gatto, R. Oyzerski, M. Moretto, N. Gutin, M. Stefanini, Y. Chen, C. Segala, C. Davenport, L. Dematte, A. Mraz, J. Battilana, K. Stormo, F. Costa, Q. Tao, A. Si-Ammour, T. Harkins, A. Lackey, C. Perbost, B. Taillon, A. Stella, V. Solovyev, J. A. Fawcett, L. Sterck, K. Vandepoele, S. M. Grando, S. Toppo, C. Moser, J. Lanchbury, R. Bogden, M. Skolnick, V. Sgaramella, S. K. Bhatnagar, P. Fontana, A. Gutin, Y. Van de Peer, F. Salamini and R. Viola. 2007. A high quality draft consensus sequence of the genome of a heterozygous grapevine variety. PLoS ONE 2 (12): e1326.

Verries, C., C. Bes, P. This and C. Tesniere. 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, C. M., R. Kalendar, K. Anamthawat-Jonsson, A. Suoniemi and A. H. Schulman. 1999. Structure, functionality, and evolution of the BARE-1 retrotransposon of barley. Genetica 107: 53-63.

Vos, P., R. Hogers, B. Marjo, M. Reijans, T. Van de Lee, M. Hornes, A. Frijters, J. Pot, J. Peleman, M. Kuiper and M. Zabeau. 1995. AFLP: a new technique for DNA fingerprinting. Nucleic Acids Res. 23: 4407-4414.

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

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

Figures

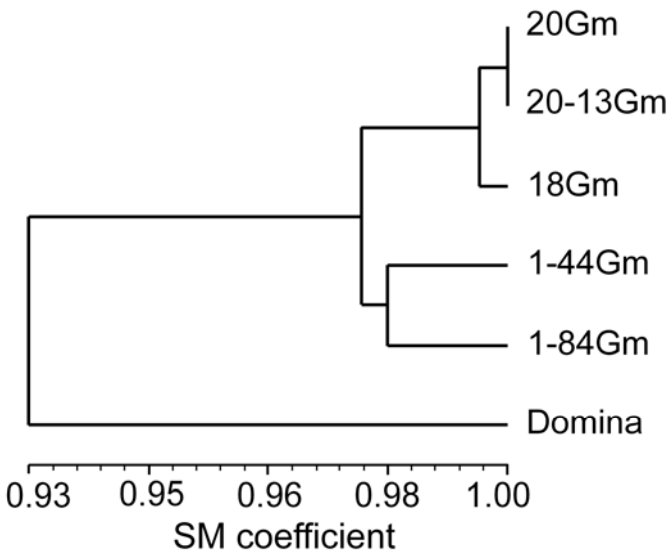

Figure 1: Gel including three primer combinations (F0100-M27, F0103-M27, F0104-M27). Number 1 to 6 is the order of clones (in repetitions): 1=20Gm, 2=1-44Gm, 3=18Gm, 4=20-13Gm, 5=1-84Gm, 6=Domina (ST49)

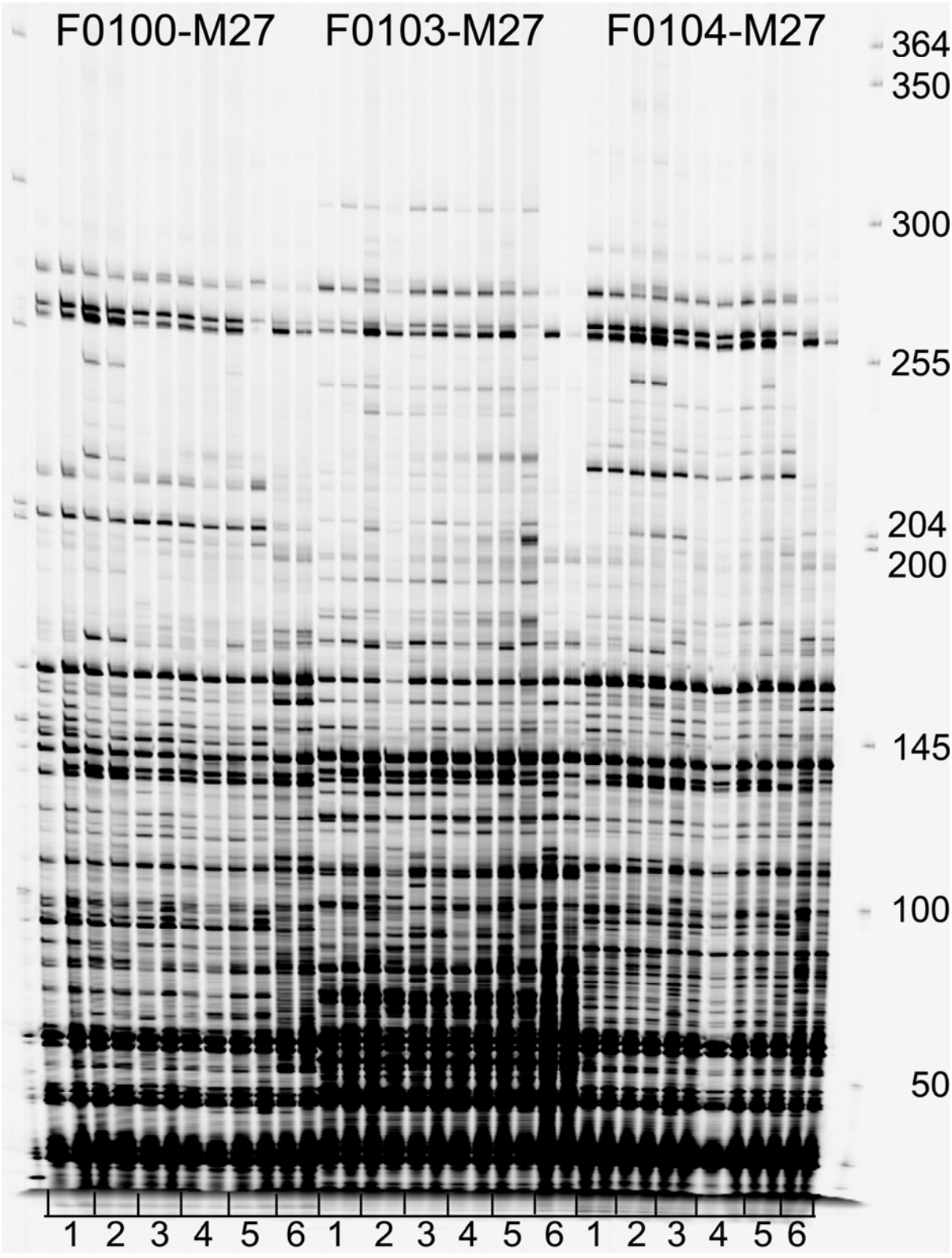

Figure 2: UPGMA based dendrogram (using the SM index) describing the genetic relationship among five Pinot noir and one Domina clones, obtained by S-SAP.

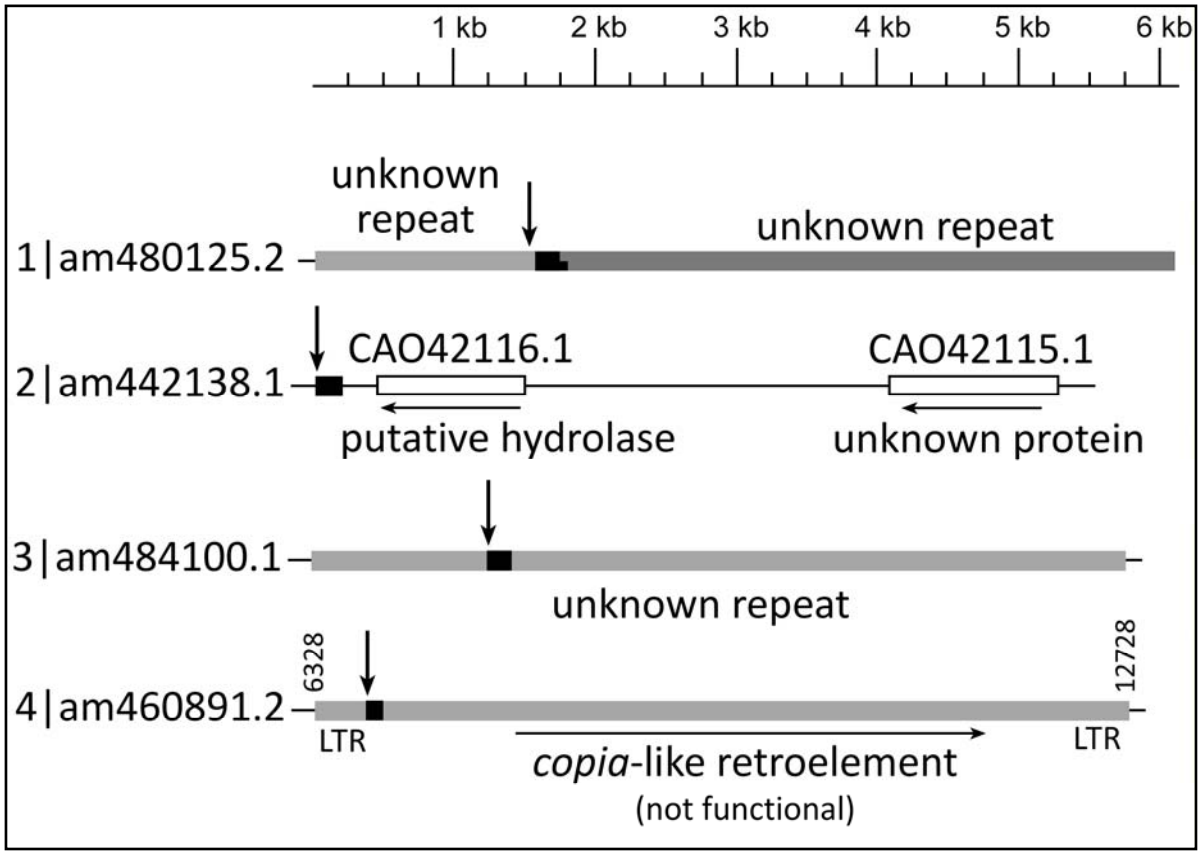

Figure 3: genomic position of the sequenced bands (in black). Accession numbers of sequences to which the band sequences match are given on the left. Coordinates of the LTR element are given for sequence 4. Grey boxes represent repeatable sequences and white boxes are putative genes above which the accession numbers of predicted proteins are given (arrow indicate the direction of the translation). Vertical arrows represent *Mse*I restriction sites for the sequenced bands. All sequences are drawn on scale.

Tables

Table 1: list of *Mse*I primers and universal retrotransposons-based primers used in the PCR analyses.

Primer Code	DNA sequence
M(0)	5’ GATGAGTCCTGAGTAA 3’
M22	5’ GATGAGTCCTGAGTAACAA 3’
M23	5’ GATGAGTCCTGAGTAACTT 3’
M24	5’ GATGAGTCCTGAGTAACAC 3’
M25	5’ GATGAGTCCTGAGTAACAT 3’
M27	5’ GATGAGTCCTGAGTAACTG 3’
F0100*	5’ TAGGTCGGAACAGGCTCTGATACCA 3’
F0103*	5’ ACCGAGCAACTTGAGCTCTGATACCA 3’
F0104*	5’ CTAGGGTCAAGGGGGCTCTGATACCA 3’
F0105*	5’ GGGAAATGGTCCGCTCTGATACCA 3’
F0113*	5’ AGTTCATCGTAGGTGGGCGCCA 3’
F0117*	5’ ATCCCCAGCGGAGTCGCCA 3’

*Revealed prior publication (Kalendar et al., submitted).

Table 2: Similarity matrix derived from S-SAP analysis.

20Gm	1-44Gm	18Gm	20-13Gm	1-84Gm	Domina
1					
0.9731343	1				
0.9910448	0.9701493	1			
0.9955224	0.9746269	0.9925373	1		
0.9791045	0.9791045	0.9761194	0.980597	1	
0.9238806	0.9358209	0.9268657	0.9283582	0.9358209	1

Table 3: Four polymorphic bands of different clones and primer combinations were sequenced and used for sequence analyses. *Mse*I primer sequences are underlined (missing in sequence 1). PBS primers are double underlined.

Information	Sequence
1[a] 1-44Gm[b] F0117 - M24[c] 244 bp[d] 0%[e]	TGAAATAGCTATGATGCTCCCAAAACTCCTGTGG AGTGCGTGGTGTGGATCTTCAAAGYGGCAGCCC CCTTTCCAAAAATTCCATGCCTGTAAGTATCTCTC CCATCTACAACCCTAAAGCATTCTCCAGCACAGT TTGCTTCATACCGGAATCATAGAACTGTGAGWTC ATGTTGTCATCAACTGAGACCTTCTCATGAACCG AGAGGCACGGGGWCCACAAAATGGCGACTCCG CTGGGGAT
2[a] 1-44Gm[b] F0104 - M24[c] 218 bp[d] 35%[e]	GATGAGTCCTGAGTAACACAACCCATTGCATCTA GTGCCAGTCGTTGTTTTAATGTATGGTGCCCGTC TCTTGTTGCTTGTTGGCTCAAAAGTAGAAATAATT ANATTTGACTCCTAGTACTATAACTTTTCCACCTT TGAAAAGGCCCCATGATGTCTCTCCTAAATGCCT AATAAGTTAGAACATCCACATGGTATCAGAGCCC CCTTGACCCTAG
3[a] 18 Gm[b] F0104 - M27[c] 221 bp[d] 60%[e]	GATGAGTCCTGAGTAACTGTTTGGAGATGGCTTG GTCCAGAAAATGCCAAAGTGCAGTCATGGGTGTT TTCATATGGCGAATGAGGGTTTGGTACAAGCCTT GCAAGCATTATAGCAGTCAGATTGATAACCCCAA GGTAGTGATCTCGGTATTGGGCACGAGCATTAGA AACAGTCCAGATCAAGTCTCACCCGTGGTATCA GAGCCCCCTTGACCCTG
4[a] 18 Gm[b] F0103 - M25[c] 156 bp[d] 23%[e]	GATGAGTCCTGAGGTAACATGGCCGCGTGTTCTT CAAACCGGTATGTAATCAATTCGTTAAATTTTTGA GATGTGTTGAATTCAATGATCTTGAATTTGTGTGT TAATTTTCRCGTTAAATGCTAACAATTGGTATCAG AGCTCAAGTTGCTCGGT

[a] Nº of sequence; [b] Clone name; [c] Primer combination; [d] Sequence length

[e] Mismatch of the PBS primers.

Table 4: Detected polymorphism by six universal primers. Total number of markers and polymorphic markers for Pinot/Domina clones.

Primer	Number of markers	Total polymorphic markers	Polymorphic markers for *V. vinifera* cv. Pinot noir clones (%)	Markers polymorphic only for *V. vinifera* cv. Domina
F0100	174	20	13 (7,5 %)	7
F0103	132	10	5 (3,8 %)	5
F0104	171	15	6 (3,5 %)	9
F0105	69	4	2 (2,8 %)	2
F0113	54	4	2 (3,7 %)	2
F0117	70	6	4 (5,7 %)	2
Total	670	59	32 (4,8 %)	27

ARTICLE 4

LTR-retrotransposons of grapevine and their implementation for the IRAP and REMAP fingerprinting

Andrej Benjak[1], Ruslan Kalendar[2], Alan H. Schulman[2] and Astrid Forneck[1]

This work was done in collaboration with Dr. Ruslan Kalendar and Prof. Dr. Alan H. Schulman at the Plant Genomics Laboratory, Institute of Biotechnology, University of Helsinki, Finland. The article presented here is a draft result report which will be soon submitted for publication in a scientific journal. The results are shown here only for educational purposes.

LTR-retrotransposons of grapevine and their implementation for the IRAP and REMAP fingerprinting

Andrej Benjak[1], Ruslan Kalendar[2], Alan H. Schulman[2] and Astrid Forneck[1]

[1] Institute of Horticulture and Viticulture, University of Natural Resources and Applied Life Sciences, Vienna, Austria.

[2] Plant Genomics Laboratory, Institute of Biotechnology, University of Helsinki, Finland.

Results and Discussion

The goal of the project was to characterize LTR-retrotransposons in the grapevine genome and use the LTR sequences to design primers suitable for the IRAP and REMAP fingerprinting methods. We used the FastPCR software (http://www.biocenter.helsinki.fi/bi/Programs/fastpcr.htm) to mine for LTR sequences in the sequenced genome of grapevine. The algorithm used by the FastPCR relies on several universal primers for plant LTR-retrotransposons. The primers were designed based on the conserved PBS sequences of LTR-retrotrasnposons and are suitable for targeting uncharacterized LTR-retrotransposons in different plant species (Kalendar *et al*., submitted). The software is searching for hits to any of the PBS primers and extracts the targeted sequences with its 5' flanking sequence (up to 150 bp) which should, in theory, correspond to the 3' of the LTR. This strategy of transposon mining was chosen because our primary goal was to retrieve LTR sequences from the sequenced genome, rather than completely analyze the full length elements. We have retrieved a total of 4,192 putative LTR sequences which were clustered into groups according to sequence similarity. The clustering was also done using the FastPCR software into which an innovative algorithm was implemented. The algorithm looks for sequence similarities through "words" of certain size (similar to the BLAST algorithm) and creates an identity matrix summarizing the number of hit windows that match a given sequence. The matrix is then grouped into clusters of sequences. The advantage of this method is that no multiple alignments are needed for the clustering; while the drawback is that the clustering is quite strict and is not based on evolutionary models of nucleotide substitutions.

We have designed 53 primers for the most abundant clusters and tested them as single primer PCR on one grapevine sample of Pinot Noir (Figure 1).

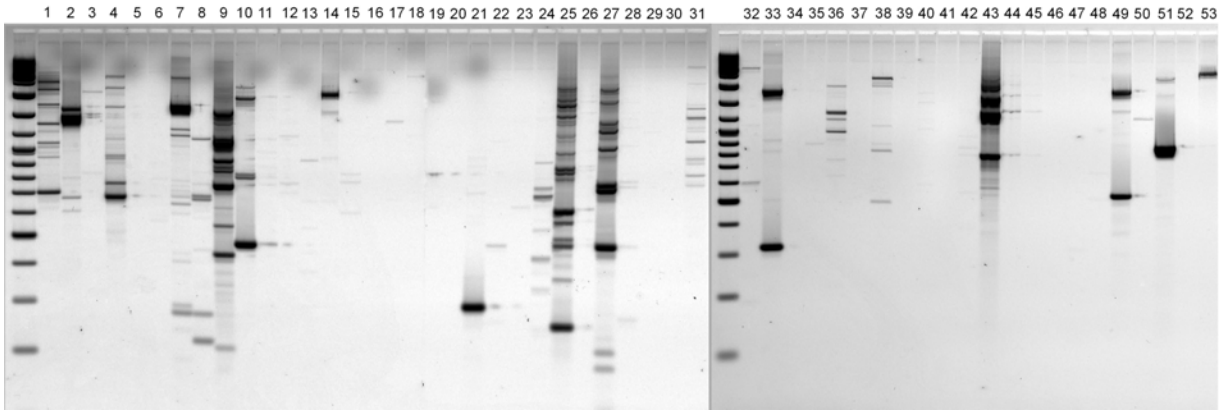

Figure 1: Single primer PCRs for 53 different LTR-based primers.

IRAP optimization

For the IRAP method, normally two primers targeting different families of LTRs are used. The number of primers in our case is too high to test all theoretical primer combinations. We chose the first 2 primers in combination with the others to test the efficiency of the method (on Pinot Noir DNA sample). Results are given in Figure 2 and 3.

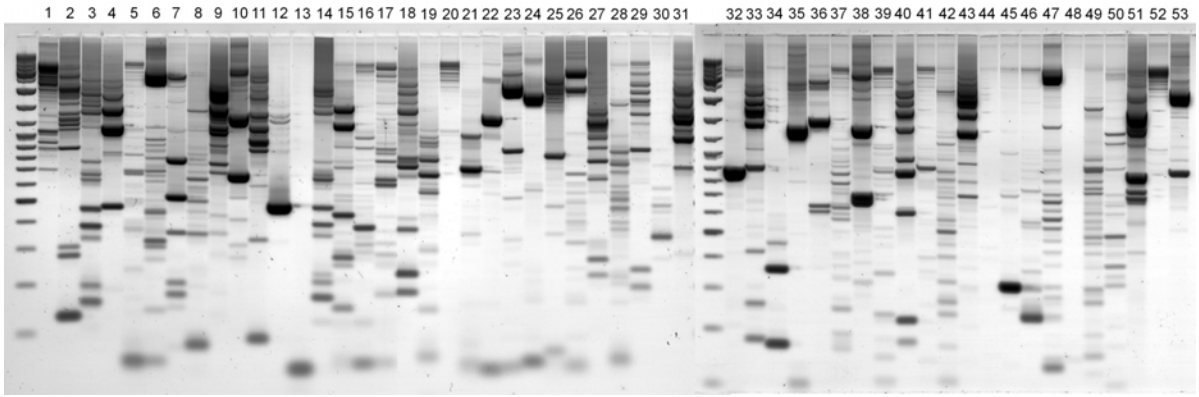

Figure 2: IRAP test with all the primers in combination with the primer 1.

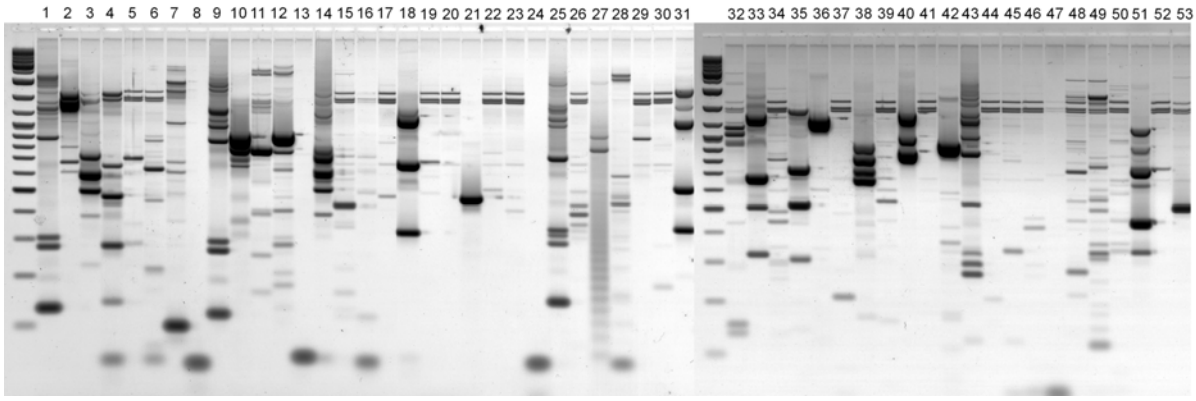

Figure 3: IRAP test with all the primers in combination with the primer 2.

More bands appeared when primer combinations were used, even with those primers that did not produce any bands in single primer PCR. In theory, bands that are produced with single primers should appear also in their combinations with other primers. In most cases some intense bands from single primer PCRs are less pronounced in primer combinations and some bands are even absent. This might be explained by the competition of the second primer for the Taq and dNTPs, as well as possible annealing between the two primers used in combination.

REMAP optimization

REMAP method combines LTR-specific primers and SSR primers. We first tested 10 SSR primers using Pinot Noir DNA (figure 4). We chose SSR primers 2-4 to combine with the LTR primers and got a banding pattern similar to that of IRAP (figure 5). Some combinations of primers produced a smear instead of bands.

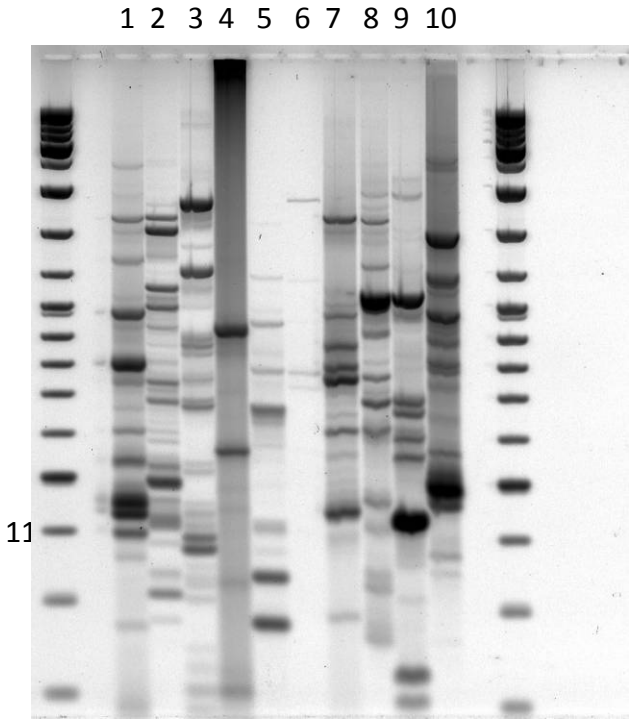

Figure 4: single SSR primer test primers:

1. $(TG)_{10}$ A
2. $(AC)_{10}$ G
3. $(AC)_{10}$ T
4. $(CT)_{10}$ G
5. $(TG)_{10}$ C
6. $(GA)_{10}$ C
7. $(CTC)_6$ G
8. $(AC)_{10}$ C
9. $(AG)_{10}$ C
10. $(AGC)_6$ C

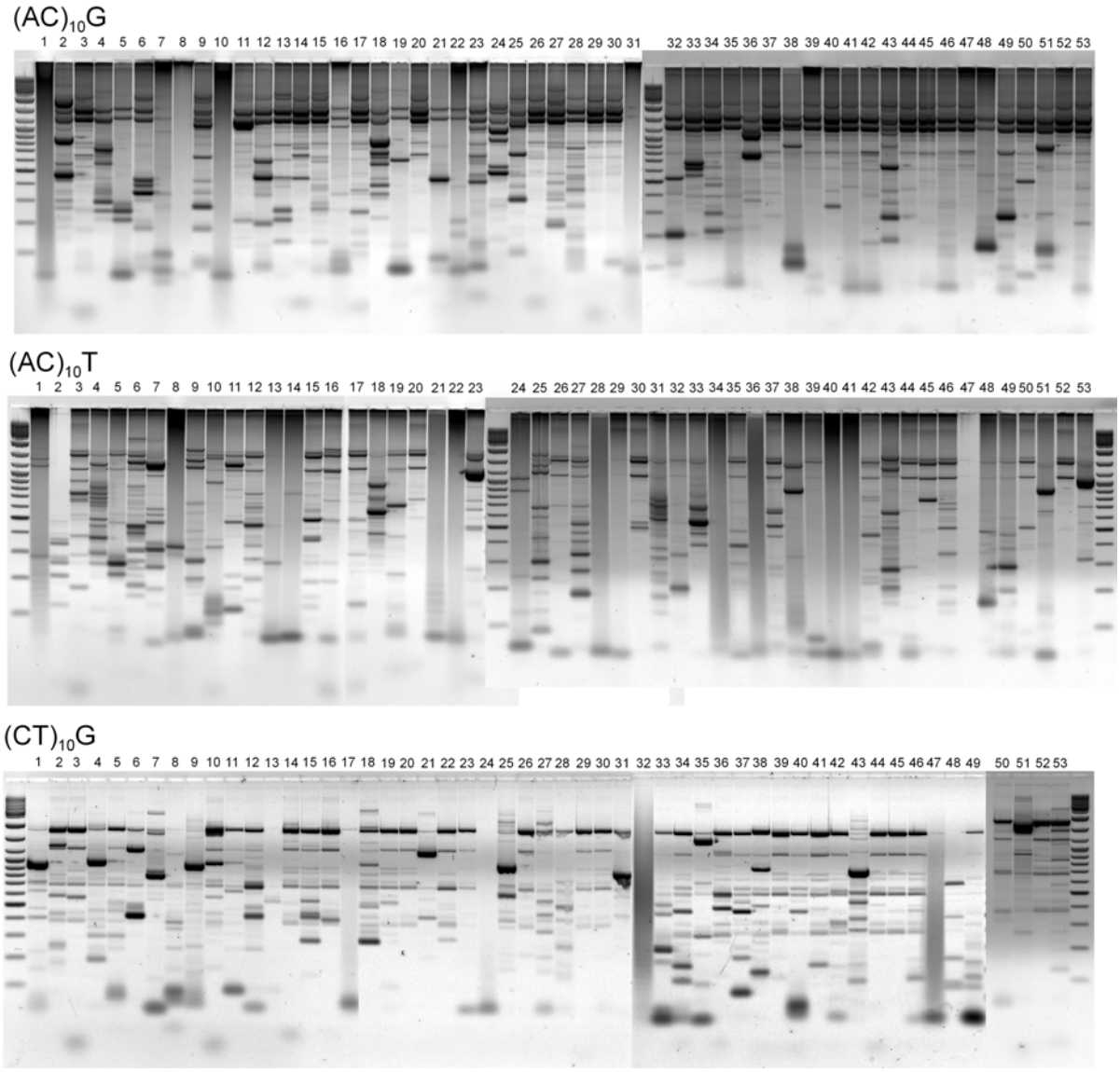

Figure 5: Three different SSR primers in combination with the 53 LTR primers.

IRAP and REMAP in grapevine fingerprinting

26 samples were extracted (Table 1) and used for the IRAP and remap fingerprinting with different primer combinations (Figures 6-8).

Table 1: Samples used for the fingerprinting analysis.

Sample №	Cultivar	Clone	Origin
1	Pinot	777	LVWO Weinsberg, Germany
2		We 242	
3		We 111	
4		20 Gm	
5		1-84 Gm	
6		18 Gm	
7	Riesling	198-44 Gm	Institute of Grapevine Breeding Geisenheim, Germany
8		64-183 Gm	
9		110-14 Gm	
10		24-195 Gm	
11		239-17 Gm	
12	R. Riesling	23 Gm	
13	Chardonnay	50 Gm	
14		1 Gm	
15		33 Gm	
16		3 Gm	
17		52 Gm	
18	Cabernet Sauvignon	Levadoux	LVWO Weinsberg, Germany
19		Gm 1	
20	Cabernet Mitos	-	Rebveredler Antes, Heppenheim, Germany
21	Cabernet Cortis	-	
22	Cabernet Dorsa	We 750	
23	Lemberger x Cab. Sau.	We 70-281-36	LVWO Weinsberg, Germany
24	Lemberger x Cab. Sau.	We 70-281-37	
25	Acolon		
26	*Vitis riparia* – Michaux	-	

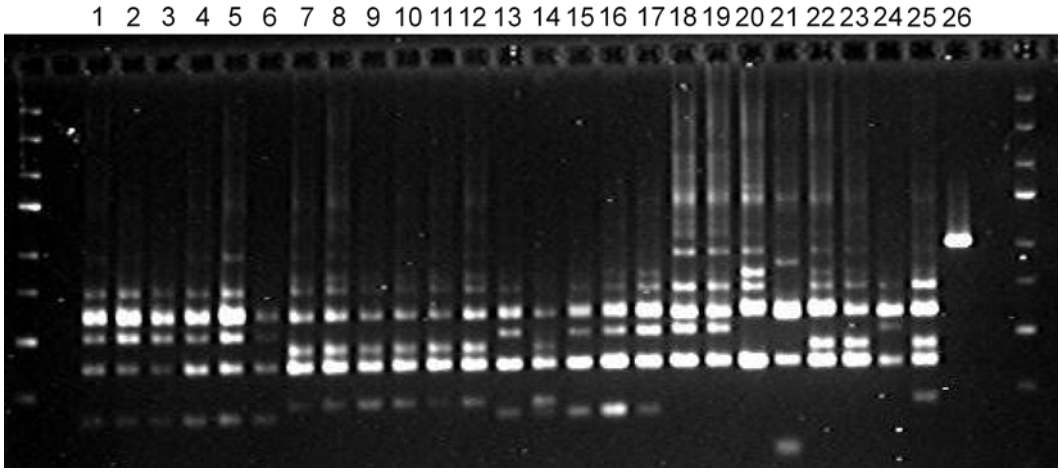

Figure 6: IRAP using primers 2 and 3 in combination. Sample information is given in Table 1.

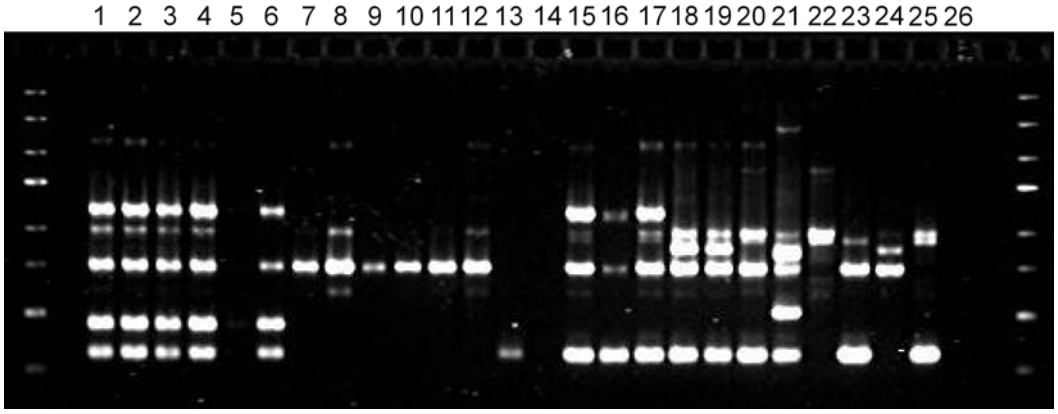

Figure 7: IRAP using primers 2 and 51 in combination. Sample information is given in Table 1.

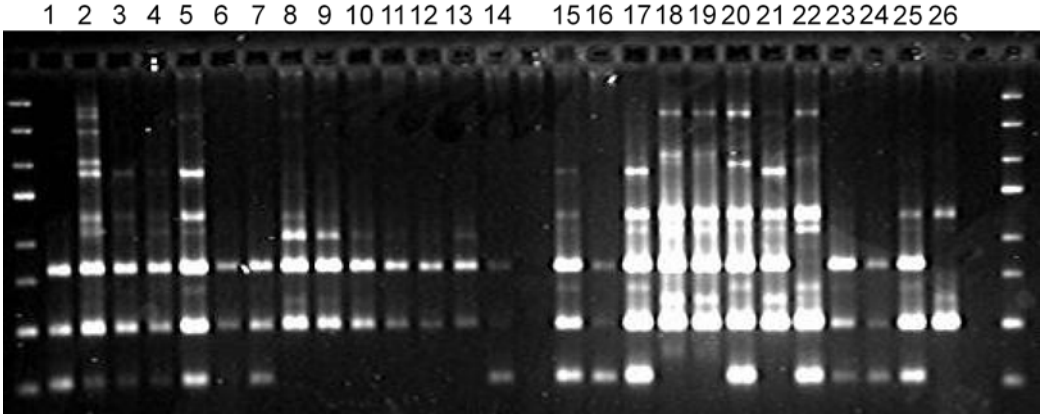

Figure 8: REMAP using SSR-primer $(CT)_{10}G$ and LTR-primer 18 in combination. Sample information is given in Table 1.

Analysis of the LTR sequences

In order to confirm that the sequences mined with the FastPCR software were indeed LTRs of retrotransposons and to have an insight into the types of elements that were targeted, all sequences from which primers were designed were analyzed more in detail. Each putative LTR sequence was blasted to the genomic sequences from NCBI to retrieve their flanking regions. Corresponding shotgun sequences were re-blasted to the database in order to spot the borders of the putative element. LTRs were searched using the "search for direct-direct repeats" option in the FastPCR software. Internal part of the putative element was blasted against the protein predictions for grapevine available at NCBI. Elements that contained the conserved domain for the reverse transcriptase (RT) were aligned according to Xiong *et al*. (1999) and a Maximum Likelihood tree was calculated from the alignments (Figure 9). Retrotransposons were classified according to sequence similarity and organization of conserved domains in the *pol* genes. Primers were blasted against shotgun sequences from Velasco *et al.* (2007) allowing 5% error (corresponds to 1-2 nucleotide mismatches) to estimate the number of primer-target sequences in the genome. Details for each cluster of sequences from which LTR-primers were designed are given in Table 2.

While most of the mined sequences were indeed LTR of different groups of retrotransposons, 2 clusters correspond to LINE elements (primer 37 and 44). We found also some TRIMs or solo LTRs, which do not contain coding capacities and were not included in the alignment. . In some cases (data not shown) a cluster sequence corresponded to the internal part of an LTR-retrotransposon. The alignment clearly groups our sequences into the major groups or retrotransposons: Gypsy, Copia and LINE. Many sequences cluster very close to each other, but further analyses confirmed that most of them represent different families of elements (data not shown). The problem is that the RT domain used for the alignment is the most conserved part of these elements and might not provide enough resolution to differentiate the elements into families, but it is useful for comparison of different groups of retrotransposons including elements from other species.

Table 2: Information on LTR-primers and their deriving sequences.

Primer	Cluster size	Primer occurrence in the genome	representative	coordinates	Classification
1	182	2362	Gypsy6-VV	In Repbase	Gypsy
2	110	859	CAAP02002665.1	13404-5812	Gypsy
3	108	1612	AM423269.2	24586-13332	Gypsy
4	91	2478	CAAP02000264.1	119593-141437	Gypsy
5	86	317	AM479628.1	882-5668	Copia
6	83	726	AM464471.1	37516-32428	Copia
7	79	509	AM423633.2	13520-5012	Gypsy
8	72	254	AM484257.1	29510-29914	TRIM
9	60	988	GYVIT1	In Repbase	Gypsy
10, 38	87	418-462	EF439837.1	Full length	Gypsy
11	53	604	AM443653.2	2020-13119	Gypsy
12	51	108	CAAP02003565.1	10071-36917	Gypsy
13	40	182	AM439905.2	11194-10479	Gypsy
14	46	1576	CAAP02003507.1	7124-20094	Gypsy
15	46	158	AM426739.2	149-5075	Copia
16	35	113	AM428732.2	42086-37022	Copia
17	29	166	AM431245.2	15859-10641	Copia
18	38	498	CAAP02002180.1	21311-12367	LTR-retrotransposon
19, 20	73	366-299	EU009622.1	complete	Copia
21	36	235	AM475140.2	5544-7419	TRIM (3x tandem LTR)
22	35	145	CAULIV1	In Repbase	Caulimovirus-like
23	34	123	AF116598.1	1193-3584	Copia
24	34	313	AM444308.1	21341-19053	LTR-retrotransposon
25	33	1748	Gypsy17-VV	In Repbase	Gypsy
26	30	58	AM480082.1	5806-11206	Copia
27	32	754	Gypsy11-VV	In Repbase	Gypsy
28	24	35	CAAP02002966.1	33756-14027	Gypsy
29, 39	39	486	Gypsy12-VV	In Repbase	Gypsy
30	23	48	EU009616	complete	Copia
31	26	2569	AM469731.2	5298-4044	putative solo LTR
32	18	268	Gypsy3-VV	In Repbase	Gypsy
33	17	206	CAAP02001158.1	54954-73599	Gypsy
34	17	103	CAAP02001971.1	49721-29824	Gypsy
35	17	104	Gypsy19-VV	In Repbase	Gypsy
36	32	303	Gypsy22-VV	In Repbase	Gypsy
37	16	25	AM475512.2	955-7151	LINE
40	15	36	CAAP02001282.1	69172-85060	Class I
41	15	44	AM462475.1	8261-1726	Copia
42	15	30	CAAP02001158.1	54954-73599	Gypsy
43	421	1615	CAAP02002456.1	40288-51666	Gypsy
44	14	23	CAAP02002910.1	22278-28609	LINE
45	12	34	CAAP02000843.1	3510-23732	Gypsy
46	12	30	CAAP02002145.1	31876-49600	LTR-retrotransposon
47	12	35	CAAP02000004.1	307852-313465	Gypsy
48	18	155	CAAP02000406.1	101710-102062	TRIM
49	18	54	CAAP02001215.1	1523-6509	Class I
50	18	67	CAAP02002020.1	19339-34525	Gypsy
51	11	85	CAAP02003050.1	40184-45936	Gypsy
52	11	89	CAAP02002332.1	35320-29531	Copia
53	22	338	CAAP02000529.1	10349-16728	LTR-retrotransposon

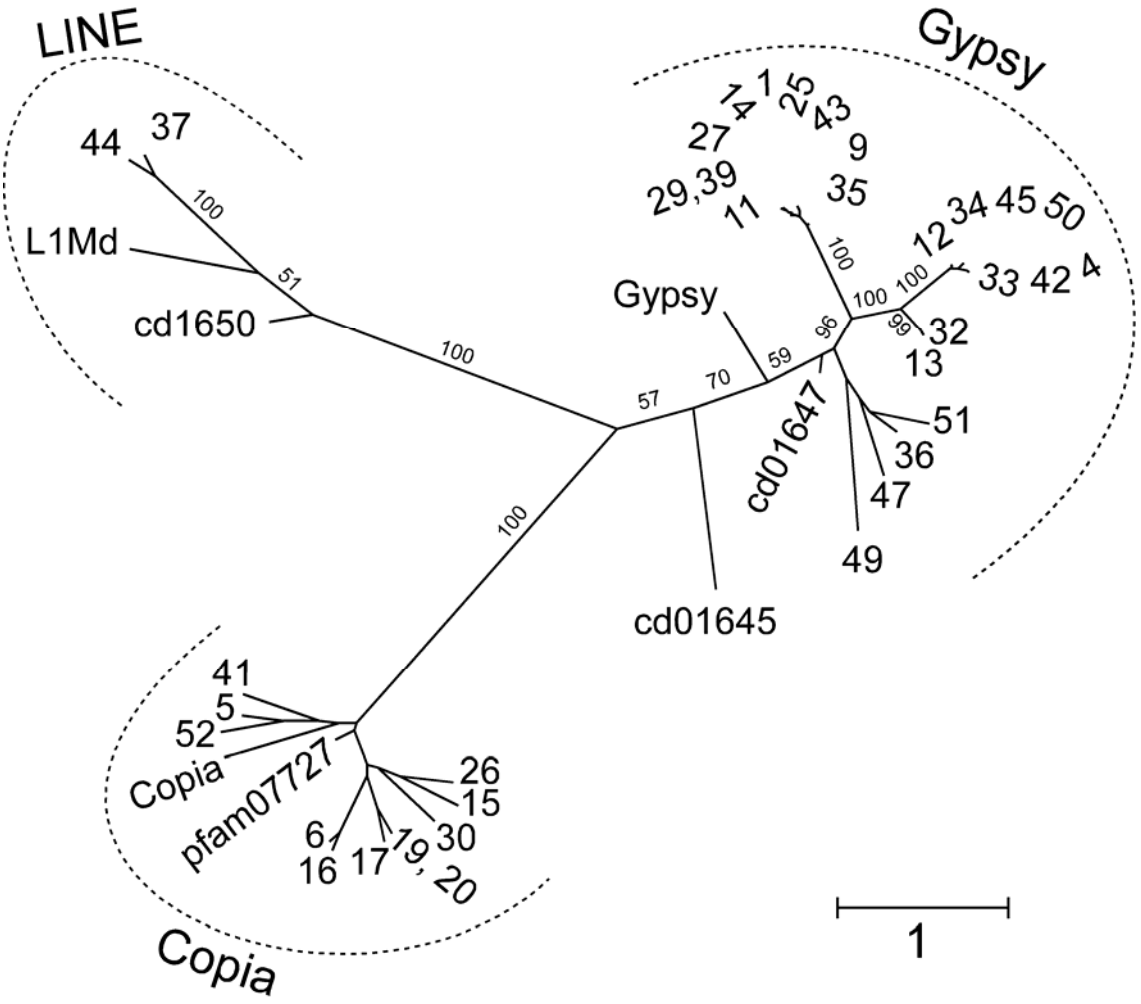

Figure 9: Maximum Likelihood tree of the alignment of the reverse transcriptase domain of the retrotransposons analyzed in this work. Included in the tree are conserved domains most similar to the sequences analyzed (pfam number) as well as Line1, Copia and Gypsy sequences as shown in the alignment in Xiong *et al*. (1999).

ARTICLE 5

Genome-Wide Analysis of the “Cut-and-Paste” Transposons of Grapevine

Andrej Benjak, Astrid Forneck, Josep M. Casacuberta

Status: Open access publication

OPEN ACCESS Freely available online

PLoS one

Genome-Wide Analysis of the "Cut-and-Paste" Transposons of Grapevine

Andrej Benjak[1,2], Astrid Forneck[2], Josep M. Casacuberta[1]*

1 Departament de Genètica Molecular Vegetal, Centre de Recerca en Agrigenòmica (CRAG), Barcelona, Spain, **2** Institute of Horticulture and Viticulture, University of Natural Resources and Applied Life Sciences, Vienna, Austria

Abstract

Background: The grapevine is a widely cultivated crop and a high number of different varieties have been selected since its domestication in the Neolithic period. Although sexual crossing has been a major driver of grapevine evolution, its vegetative propagation enhanced the impact of somatic mutations and has been important for grapevine diversity. Transposable elements are known to be major contributors to genome variability and, in particular, to somatic mutations. Thus, transposable elements have probably played a major role in grapevine domestication and evolution. The recent publication of the complete grapevine genome opens the possibility for an in deep analysis of its transposon content.

Principal Findings: We present here a detailed analysis of the "cut-and-paste" class II transposons present in the genome of grapevine. We characterized 1160 potentially complete grapevine transposons as well as 2086 defective copies. We report on the structure of each element, their potentiality to encode a functional transposase, and the existence of matching ESTs that could suggest their transcription.

Conclusions: Our results show that these elements have transduplicated and amplified cellular sequences and some of them have been domesticated and probably fulfill cellular functions. In addition, we provide evidences that the mobility of these elements has contributed to the genomic variability of this species.

Citation: Benjak A, Forneck A, Casacuberta JM (2008) Genome-Wide Analysis of the "Cut-and-Paste" Transposons of Grapevine. PLoS ONE 3(9): e3107. doi:10.1371/journal.pone.0003107

Editor: Edward Newbigin, University of Melbourne, Australia

Received June 19, 2008; **Accepted** August 10, 2008; **Published** September 3, 2008

Funding: This work was partially funded by a grant of the Ministerio de Educación y Ciencia (Grant BFU2006-04005) and the Xarxa de Referència en Biotecnologia from the Generalitat de Catalunya to J.M.C. A.B. was partially funded by an EMBO Short Term Fellowship (ASTF 115-07). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: jcsgmp@cid.csic.es

Introduction

The grapevine (*Vitis vinifera* L.) is a widely cultivated crop that has accompanied the development of human culture since its domestication in the Neolithic period (c. 8500-4000 BC). Cultivated grapevine (*Vitis vinifera* spp. *sativa*) is supposed to have been domesticated from wild grapevine populations (*Vitis vinifera* spp. *sylvestris* Gmelin) in the Near East, from where its culture expanded through Europe [1], although recent results suggest that different domestication events took place in both East and West Europe [2,3]. The domestication of grapevine has undergone a selection for traits important for its cultivation and usage (e.g. vigor, hermaphrodite flowers, berry content and size, cluster structure). Although sexual crossing has been a major driver of grapevine evolution, its vegetative propagation enhanced the impact of somatic mutations and has been important for grapevine diversity. Clonal selection of superior individuals identified by growers has led to many clones with different phenotypes while maintaining the same cultivar [4]. Some of these mutations exist and are maintained in a chimeric state affecting only single cell layers [5], the phenotype of the plant being the result of the combination in different cells of two different genotypes.

Transposable elements (TEs) are known to be major contributors to genome variability and, in particular, to somatic mutations. Plant genomes contain high albeit variable amounts of TEs that account for 15–80% of their genome. Most plant TEs are activated in somatic cells by different biotic and abiotic stresses including wounding, and they are usually silent in germinal cells, which limits their mutagenic capacity and their ability to colonize plant genomes (e.g. [6]). The propagation of grapevine includes layering (in the native habitats), cutting of dormant and green shoots, grafting and sometimes tissue culture steps. This practice enhances the impacts of somatic mutations and possibly increases the chance of TEs to transpose and multiplicate. Thus, TEs could have been a major force creating the variability used for grapevine breeding from its domestication to present times. Indeed, the skin color in white grapes, a highly desired trait for grape berry and wine quality, has been shown to be the consequence of a retrotransposon insertion in the promoter of a *Myb*-related gene that regulates anthocyanin biosynthesis [7]. This mutation is present in most white grape varieties [8,9].

Transposable elements are usually classified in two major groups based on their structure and transposition mechanism: Retrotransposons or class I elements, which transpose by an RNA intermediate, and class II or DNA transposons, which use an intermediate of DNA. Up to now, in addition to *Gret1*, the element responsible for the grape color phenotype, two other retrotransposons have been characterized in grapevine [10,11]. On the

contrary, although there is a handful of sequences of grapevine class II elements deposited in the Repbase database (www.girinst.org) up to now no DNA transposon has been characterized in detail in this plant.

Recently, two articles describing the *Vitis* genome have been published [12,13] and shotgun sequences of grapevine genome have been made available opening the possibility for a genome-wide bioinformatical analysis. We present here a global and detailed analysis of the "cut-and-paste" class II transposons present in the genome of *Vitis vinifera* L. We characterized 1160 potentially complete grapevine transposons as well as 2086 defective copies. Our results show that these elements have transduplicated and amplified cellular sequences and some of them have probably been domesticated (i.e. have lost their ability to transpose and fulfill cellular functions, as a conventional cellular gene). In addition, we provide evidences of recent mobility of some of these elements showing the high mutagenic capacity of grapevine transposons and their capacity to induce genomic variability in this species.

Results and Discussion

The "cut-and-paste" transposon landscape in *Vitis vinifera*

Most class II transposons excise from the donor site as double-stranded DNA which is reinserted elsewhere in the genome by a mechanism usually known as "cut-and-paste" transposition. The only class II elements that transpose by a different mechanism are *Helitrons* and related elements, that transpose by rolling-circle replication, *Mavericks*, whose transposition mechanism is not yet known [14], and the bacterial *IS200/605* family of insertion sequences that transpose as a single stranded transposon circle [15,16]. "Cut-and-paste" class II transposons typically contain terminal inverted repeats (TIRs) and encode a transposase that catalyses their mobilization. The sequence and structure of the transposase together with the sequence of the TIRs recognized by this protein and the characteristics of the flanking target site duplication generated by the transposase upon inserting the element has been used to classify class II elements in ten different superfamilies: *CACTA*, *hAT*, *Merlin*, *Mutator*, *P element*, *PIF*, *piggyBac*, *Tc1/Mariner*, *Transib* and *Banshee* [14,17,18]. In plants, only elements belonging to the *CACTA*, *hAT*, *Mutator*, *PIF*, and *Tc1/Mariner* superfamilies have been described to date [14].

We searched the grapevine genome sequence for the presence of class II transposons of the five superfamilies by means of blastx searches of the shotgun sequences made publicly available by Velasco et al. [13] and using the sequences made available later by Jaillon et al. [12] for confirmation (see Materials and Methods section for details). We have not been able to detect any grapevine sequence that could represent a *Tc1-Mariner* element. Although few sequences with very limited similarity (below the threshold set) to these elements exist, they probably represent old defective elements and were not included in this analysis. We found representatives of the other superfamilies of elements: *CACTA*, *hAT*, *Mutator*, *PIF*. We have characterized a total of 1160 potentially complete DNA transposons, as well as 2086 defective elements, which altogether represent 1.98% of the *Vitis* genome (Table 1).

The two recent reports on the draft sequence of the genome of *Vitis vinifera* spp. *sativa* contain a general analysis giving an overview of the transposon content in this genome [12,13]. Both reports predict higher copy numbers of DNA-transposon-related sequences (6,344 and 9,562 respectively) compared to our results, but with substantially lower transposon content in terms of genome fraction (0.43% and 1.6% respectively). The reported mean length of the described copies is low (0.3 Kb/element and 0.9 Kb/element respectively), possibly because the characterized sequences are limited to the well conserved coding regions of TEs and thus miss most of the transposon sequences which are non-coding. We have performed a stringent search and have characterized these elements in their full sequence (up to the TIRs when present) omitting only TEs deleted copies representing less than 20% of the length of the complete TE representative for each family. Employing these parameters for analysis is crucial to research the structure and possible mobility of TEs, and analyze their capacity to transduplicate sequences or become domesticated. Our analysis shows the mean TE length of 3.3 Kb/element, which is more than three times bigger when compared with previous reports.

In order to get insight on the evolutionary dynamics of class II TEs in grapevine we conducted a detailed TE analysis: For each superfamily we have compared the protein sequence of the putative transposase of all elements containing a transposase conserved region characteristic of this superfamily (see Methods for details). Maximum likelihood trees were generated from protein sequence alignments which allowed us to define different families for each transposon superfamily. We have analyzed the presence of STOP codons and frameshifts in the potential ORFs as well as the existence of ESTs in the grapevine databases that could suggest transcription of transposases and possible transpositional activity. Defective elements were identified for each family by blastn analyses using representatives of complete TEs as queries.

hAT is the most prevalent superfamily of transposons in grapevine

We have found 1459 *hAT*-related elements in the grapevine genome, which makes *hATs* as the most prevalent "cut-and-paste" transposon family in grapevine in terms of copy number (Table 1). The phylogenetic analysis of these elements showed that they can be grouped in different families (Figure 1 and Table 2 and Dataset

Table 1. Total number and genome coverage of class II elements in *Vitis vinifera*.

Superfamily	Copies	N° of full length copies[1]	N° of deleted copies	Mb	Coverage
hAT	1459	597	862	3.64	0.66%
PIF	236	93	143	0.6	0.11%
Mutator	1172	331	841	4.73	0.86%
CACTA	364	124	240	1.9	0.34%
Total	3231	1145[2]	2086	10.87	1.98%

[1]These are copies which have at least 90% of the putative transposase gene and represent potential full length elements (see Materials and Methods for details).
[2]Domesticated TEs were not included (15 in total).
doi:10.1371/journal.pone.0003107.t001

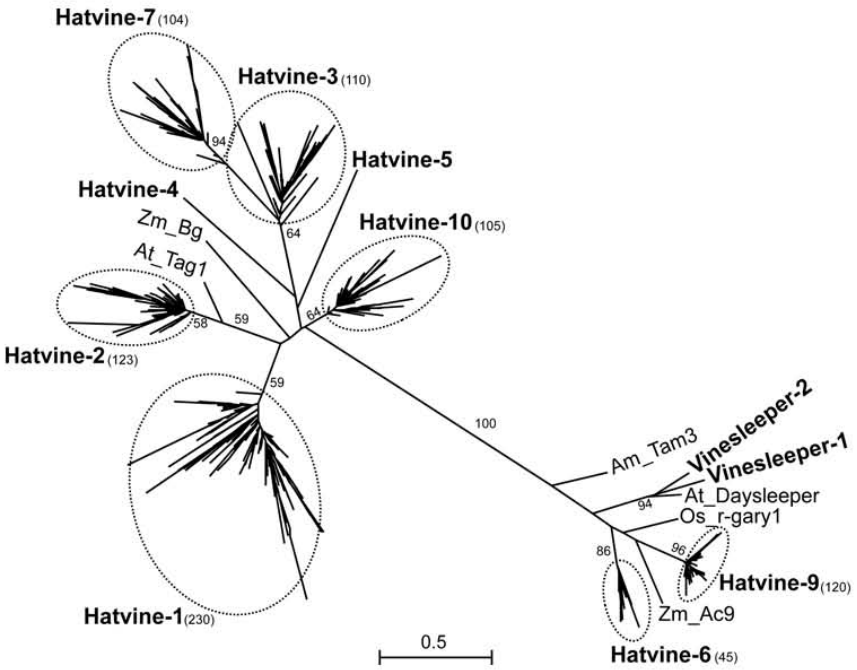

Figure 1. Maximum likelihood tree of the *hAT* superfamily. Bootstrap values higher than 50 are shown. Numbers in brackets show the number of sequences analyzed for each family. Names written in bold are *Vitis* families. Names in plain text are *hAT* elements from other plants with the first two letters representing the species name (Am = *Antirrhinum majus*, At = *Arabidopsis thaliana*, Os = *Oryza sativa*, Zm = *Zea mays*). *DAYSLEEPER* and *r-gary1* are domesticated *hAT*-related transposases.
doi:10.1371/journal.pone.0003107.g001

S1). Most of these families include a high copy number of both potentially complete and defective elements. Single copy elements were found as well. These elements possibly represent domesticated transposases and are discussed in a separate chapter (see below). The *hAT* elements belonging to the high copy number families contain TIRs of 8–23 bp, with sequences similar to that of typical *hATs* [19], and are flanked by TSDs of 8 bp, as expected for elements of this superfamily [19]. The *hAT* superfamily is relatively ancient and is widespread in eukaryote genomes [19]. Thus, the high variability of grapevine *hATs*, and the high proportion of defective elements is not unexpected. However, our results show that some grapevine *hAT* families contain potentially complete elements with the capacity to encode a transposase (Table 2), suggesting that some *hATs* could have maintained the capacity to transpose. This is the case of *Hatvine-1*, *Hatvine-2*, *Hatvine-7*, *Hatvine-9* and *Hatvine-10* families that contain a high number of potentially complete elements with intact ORFs and match to transcripts in the grapevine EST collections (Table 2).

CACTA is the less active superfamily of transposons in grapevine

CACTA elements are the most abundant class II elements in *Brassica oleracea* [20] and also seem to be highly abundant in *Triticum* [21] while they are much less abundant in *Arabidopsis* [20] where they have been found almost exclusively in pericentromeric regions [22]. In grapevine we have found only 364 *CACTA* elements, one third of which are potentially complete (Table 3 and Dataset S2). However, as grapevine *CACTAs* are very long (ranging from 10 to 25 Kb) these elements account for a significant fraction of the grapevine genome (0.34%). The high diversity of the *CACTA* superfamily in grapevine, which can be divided in at least nine different families, and the low number of elements having an intact transposase-encoding ORF, suggests that grapevine *CACTA* are relatively old elements, and most of them are probably defective. Moreover, grapevine databases contain a low number of EST sequences corresponding to the *CACTA* elements described here, suggesting that most of them are probably silent at present. Of the nine *CACTA* families only *Cactavine-2*, *Cactavine-5* and *Cactavine-13* seem to have retained the capacity to be transcribed (Table 3). Interestingly these subfamilies are phylogenetically related and may have arisen recently during grapevine evolution (Figure 2).

Grapevine contains elements of the three major MULE families *MuDR*, *Jittery* and *Hop*

The *Mutator* superfamily (named after the *Mutator* (*Mu*) element in maize [23]) is a highly abundant and diverse superfamily of class II elements in plants [24]. Elements belonging to the *Mutator* superfamily are generally called *Mutator*-like elements (MULEs). They are the most abundant transposons in many plant genomes such as *Arabidopsis thaliana* [20], *Lotus japonicus* [25] and *Oryza sativa* [26,27]. While most autonomous MULEs encode a protein similar to the MURA transposase of the *MuDR* transposon (the autonomous version of the maize *Mu* element), two other families of MULEs distantly related to *MuDR* have been recently reported in plants. The *Jittery* family described in maize [28] and shown later to be present also in other plants [25] and a family related to the fungal *Hop* element [29] which in plants has so far only been found in legumes [25]. As the three subfamilies are only distantly related we have performed an independent search for *MuDR*-like elements and for elements related to the *Jittery* and *Hop* subfamilies. A high

Table 2. List of *hAT*-related families of transposons characterized in *Vitis vinifera*.

Family name	Length of complete TE (kb)	N° of TEs having >90% TPase	N° of TEs with potentially functional ORFs	N° of deleted copies	TIR length in bp	TSD length in bp	N° of EST hits	Representative	Coordinates
Vinesleeper-1	2.1	1	1	0	-	-	4	am486739.1	9525-7456
Vinesleeper-2	2	1	1	0	-	-	11	am487463.2	4039-6070
Hatvine-1	5.5	125	7	301	18	8	9	*VIHAT1*	Repbase
Hatvine-2	4.5	90	19	68	23	8	15	*VIHAT2*	Repbase
Hatvine-3	3.9	82	28	106	16	8	3	*VIHAT3*	Repbase
Hatvine-4	2.8	1	?	0	-	-	0	am480519.1	5243-8086
Hatvine-5	4.8	1	1	0	6	7	4	am478512.2	7540-5033
Hatvine-6	variable	35	17	59	13	9	2	*hAT-6_VV*	Repbase
Hatvine-7	3.9	88	10	56	17	8	15	*hAT-7_VV*	Repbase
*Hatvine-8**	2.4	1	0	0	-	-	0	am448381.1	3245-5709
Hatvine-9	2.9	76	6	113	8	8	7	am463419.2	7518-10707
Hatvine-10	5.5	67	9	94	11	8	7	*hAT-10_VV*	Repbase
Hatvine-11	3.4	31	0	65	11	-	0	*hAT-11N_VV*	Repbase

*Hatvine-8 was not included in the phylogenetical analysis because it lacks the conserved domain used for the alignments (see Materials and Methods).

doi:10.1371/journal.pone.0003107.t002

Table 3. List of *CACTA*-related families of transposons characterized in *Vitis vinifera*.

Family name	Length of complete TE (kb)	N° of TEs having >90% TPase	N° of TEs with potentially functional ORFs	N° of deleted copies	TIR length in bp	TSD length in bp	N° of EST hits	Representative	Coordinates
Cactovine-1	13.4	30	0	45	8	-	0	*EnSpm1_VV*	Repbase
Cactovine-2	14.4	17	2	18	5	-	6	*EnSpm2_VV*	Repbase
Cactovine-3	15	13	0	8	5	3	0	*EnSpm-3_VV*	Repbase
Cactovine-4	11.4	18	1	101	6	3	0	*EnSpm-4_VV*	Repbase
Cactovine-5	21–25	14	2	7	23	3	4	*EnSpm-5_VV*	Repbase
Cactovine-6	13.8	5	0	8	13	3	0	*EnSpm-6_VV*	Repbase
Cactovine-7	?	2	0	1	10	-	0	am424884.1	1597-26953
Cactovine-8	10.5	1	0	11	-	-	0	*EnSpm-8N_VV*	Repbase
Cactovine-9	~4	0	0	5	-	-	0	CAAP02001186.1	58559-52598
Cactovine-10	~5	0	0	2	-	-	0	am460863.1	9708-4784
Cactovine-11	~4	0	0	2	-	-	0	am469125.1	155-3279
Cactovine-12	?	0	0	1	-	-	0	am480617.1	375-1681
Cactovine-13	12.7	24	2	31	5	-	4	*EnSpm-13_VV*	Repbase

doi:10.1371/journal.pone.0003107.t003

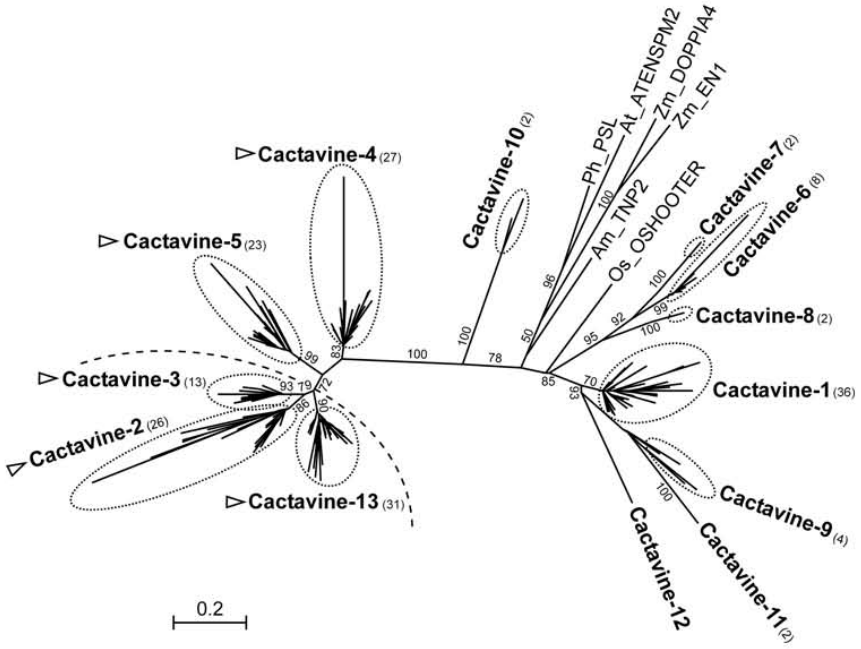

Figure 2. Maximum likelihood tree of the *CACTA* superfamily. Bootstrap values higher than 50 are shown. Numbers in brackets show the number of sequences analyzed for each family. Dashed line shows a clade of elements sharing a high similarity of the transposase gene among different families. Names written in bold are *Vitis* families. Families containing an ULP1-like region are labeled with a triangle. Names in plain text are CACTA elements from other plants taken from Repbase or NCBI with the first two letters representing the species name (Am = *Antirrhinum majus*, At = *Arabidopsis thaliana*, Os = *Oryza sativa*, Ph = *Petunia×hybrida*, Zm = *Zea mays*).
doi:10.1371/journal.pone.0003107.g002

number of MULEs related to the three families, *Mutator* (*MuDR*), *Jittery* and *Hop* were identified (Table 4 and Dataset S3).

We have characterized a total of 1172 MULEs belonging to high copy number families, 30% probably corresponding to full-length elements (Figure 3 and Table 4). Most *MuDR*-like elements belonging to the high copy number families lack an intact transposase-encoding ORF and very few of them are represented in the grapevine EST collections (Table 4), suggesting that they are old elements that mostly have lost the capacity to transpose. The *Mutavine-1* and *Mutavine-17* families could be exceptions as judged by the number of ESTs corresponding to these elements found in the grapevine databases and the existence of several elements with conserved transposase ORFs (Table 4). We have only been able to find the TSDs for a subset of MULEs, probably because of the older age of grapevine MULE insertions. However when present the TSD are always of 9 nt which is typical for MULEs in other plant genomes. Typically, MULEs have long TIRs, although a fraction of them do not [30,31]. 40% of the MULEs reported here (*Mutavine-5*, *Mutavine-6*, *Mutavine-11*, *Mutavine-13*, *Mutavine-14* and *Mutavine-17* families) do not contain TIRs, which is similar to what has been reported for *Arabidopsis* where one third of the MULEs are devoid of TIRs [30,31]. Some of these MULE families are relatively old, and the absence of recognizable TIRs could simply be due to the effect of mutations. Nevertheless in some cases, like for the *Mutavine-6* family, clear 9 nt-long TSDs were found, suggesting that these elements were mobilized in spite of their absence of TIRs, confirming the evidence found in *Arabidopsis* that non-TIR MULEs could be mobile [31]. It is interesting to note that the grapevine non-TIR MULE families do not form a monophyletic branch in a transposase-based tree (Figure 3A), suggesting a different phylogenetic history of the transposase-encoding sequences and the TIRs. This stresses the enormous variability of MULEs and their particular evolutionary dynamics [24].

In addition to the *MuDR*-like MULEs, we have found two multi-copy families of the MULEs phylogenetically related to *Jittery*-like elements and one multi-copy family, *Hopvine-1*, phylogenetically related to *Hop*, (Figure 3B). While *Jittery* elements have been found to be present in various plant genomes, up to now *Hop*-like transposons were found only in fungi and in legumes, and it has been proposed that they may have arisen during the emergence of the legume family through an ancient horizontal transfer event between fungus and legume ancestor [25]. Our results show that the *Hop* family of MULEs is more widely distributed in plants than previously thought and suggest that if these elements have been introduced into plants by fungal infections, these would have occurred several times in the evolution and would affected different plant genera. Alternatively, *Hop* elements may be an old family in plants that has been lost in most genomes except in legumes and some other species like *Vitis vinifera*. The fact that none of the 9 copies of *Hopvine-1* contains an uninterrupted ORF potentially coding for a transposase and that we have not detected any corresponding EST in the grapevine databases suggest that these elements are relatively old and have lost their capacity to be expressed and to transpose. On the contrary, the two *Jittery*-like families here characterized *Jittvine-1* and *Jittvine-2*, are expressed and could have maintained their capacity to transpose. Both families (particularly *Jittvine-1*) contain elements potentially coding for a transposase and the grapevine databases contain several ESTs that could correspond to these elements (Table 4).

Table 4. List of *Mutator*-related families of transposons characterized in *Vitis vinifera*.

Family name	Length of complete TE (kb)	N° of TEs having >90% TPase	N° of TEs with potentially functional ORFs	N° of deleted copies	TIR length in bp	TSD length in bp	N° of EST hits	representative	coordinates
MUGvine-1	1.8	1	1	0	-	-	9	am430496.2	6403-8211
MUGvine-2	2.1	1	1	0	-	-	6	am482126.1	3329-5578
MUGvine-3	1.7	1	1	1	-	-	7	am460323.1	5496-3745
MUGvine-4	2.2	1	1	0	-	-	7	am480719.1	30558-28296
MUGvine-5	1.1–1.7	2	1	0	-	-	6	am472189.1	8046-6928
MUGvine-6	2.5	1	1	0	-	-	8	am459930.1	8292-5740
MUGvine-7	2.3	1	1	0	-	-	4	am461949.2	94006-91664
MUGvine-8	2.9	1	1	0	-	-	5	am425404.1	12677-9702
Mutavine-1	17.7 kb	43	12	38	70	-	5	*Mudravi1*	Repbase
Mutavine-2	11	28	4	75	180	-	9	*Mudravi2*	Repbase
Mutavine-3	9.2–9.4	4*	0	58	158	-	0	*MuDR-3_VV*	Repbase
Mutavine-4	7.1	8*	?	57	141–144	9	0	*MuDR-4_VV*	Repbase
Mutavine-5	4.5	6	0	35	-	9	1	*MuDR-5_VV*	Repbase
Mutavine-6	10	20	4	94	-	9	1	*MuDR-6_VV*	Repbase
Mutavine-7	5.8	5	2	9	710	9	0	*MuDR-7_VV*	Repbase
Mutavine-8	9–10	26	1	72	80	-	2	*MuDR-8_VV*	Repbase
Mutavine-9	7	33	3	129	78	9	1	*MuDR-9_VV*	Repbase
Mutavine-10	2.5	1	1	0	-	-	0	am455011.1	8702-6234
Mutavine-11	4	19	0	12	-	-	0	*MuDR-11N_VV*	Repbase
Mutavine-12	10	9	5	62	416–441	9	2	*MuDR-12_VV*	Repbase
Mutavine-13	8–9	38	3	32	-	-	0	*MuDR-13_VV*	Repbase
Mutavine-14	9	24	4	50	-	-	0	am426759.2	12676-3445
Mutavine-15	?	1	1	0	-	-	3	am458922.1	4659-2101
Mutavine-16	1.8	1	1	0	-	-	0	am471827.2	3011-1176
Mutavine-17	~10 kb	6*	4	37	-	-	15	am434092.2	2041-10910
Mutavine-18	16.2	4*	0	11	230–263	-	3	am425680.2	3913-20102
Hopvine-1	4.1	2	0	7	-	-	0	am471191.1	34-4066
Hopvine-2	2.5	1	1	-	-	-	1	am457042.1	7944-5360
Jitvine-1	11.9	16	2	20	-	-	7	*MuDR-21_VV*	Repbase
Jitvine-2	14.4	35	7	42	-	-	3	am427034.2	15944-1482
Jithouse-1	4.8	1	1	-	-	-	0	am484711.1	33032-35173
Jithouse-2	2.3	1	1	-	-	-	4	am431471.2	22110-24383
Jithouse-3	2.2	1	1	-	-	-	1	am467237.2	1900-8454
Jithouse-4	2.5	1	1	-	-	-	1	am425354.2	46053-43540
Jithouse-5	2.2	1	1	-	-	-	0	am465780.1	5208-7475

*=including the ULP-1 region.

doi:10.1371/journal.pone.0003107.t004

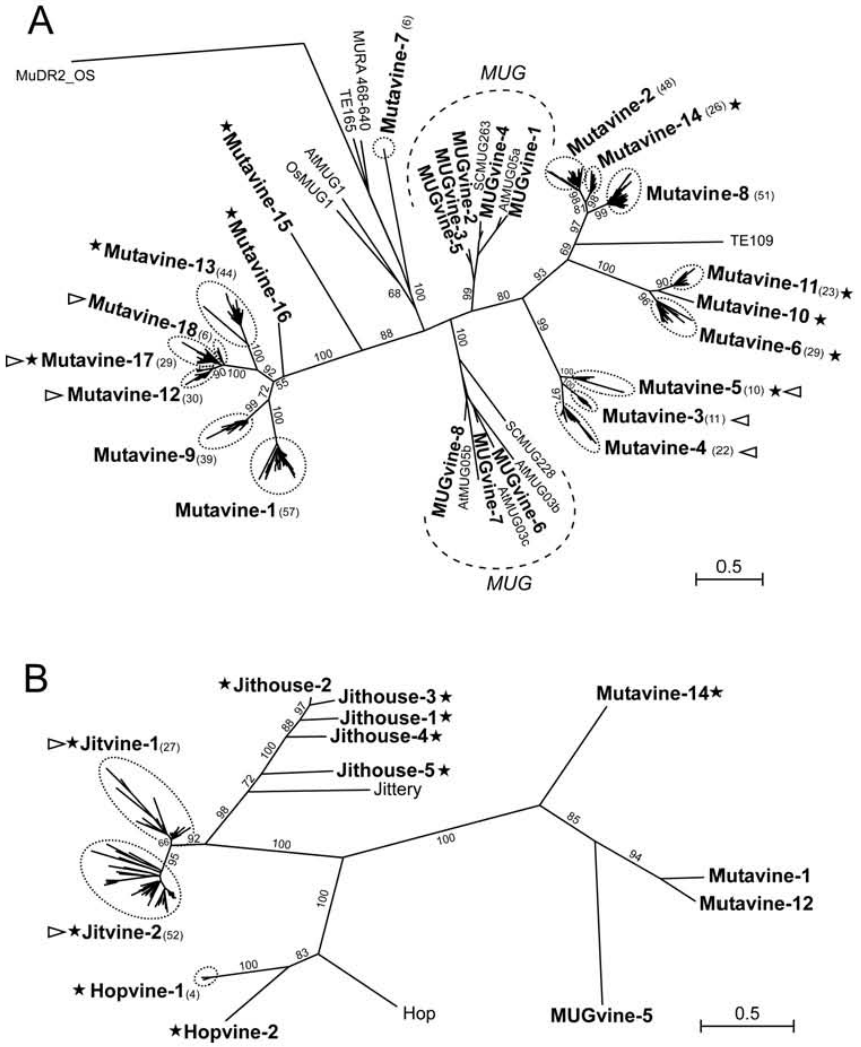

Figure 3. Maximum likelihood tree of the *Mutator* superfamily. Bootstrap values higher than 50 are shown. Numbers in brackets show the number of sequences analyzed for each family. Names written in bold are *Vitis* families. Names in plain text are *Mutator* elements from other plants (see Materials and Methods for details). Dashed lines represent domesticated *mudrA* transposases (*MUG* genes). Families in which no TIRs were found are labeled with black stars. Families containing an ULP1-like region are labeled with a triangle (pointing right for ULP1 orientated in the same frame as the TPase, pointing to the left for the opposite orientation). "A" represents all the *MuDR*-like families characterized in *Vitis* and "B" includes including the *Jittery*-like and *Hop*-like families with additional *MuDR*-like families for comparison.
doi:10.1371/journal.pone.0003107.g003

Grapevine contains potentially active *PIF* but not *Pong* elements

We have found a total of 236 *PIF/Pong*-related sequences in the grapevine genome. *Pong* elements have been shown to have undergone recent amplification in *Arabidopsis* and to a higher extend in *Brassica oleracea* whereas *PIF* elements have not been significantly amplified in both genomes [20]. The opposite was found in the genome of grapevine: *PIF* elements have attained a moderate copy number while no *Pong* element has been maintained in this genome (Figure 4). The analysis of the 236 grapevine *PIFs* shows that 93 of these elements are potentially complete, 24 of which have intact ORFs (Table 1 and 5; Dataset S4), which is the highest proportion of intact ORFs among all superfamilies analyzed in our study and strongly indicates that *PIF* elements have amplified recently during grapevine evolution. The phylogenetic analysis show that the grapevine *PIFs* group into four families and do not plot together to the families previously defined in other plant genomes [32] (Figure 4). This confirms a recent grapevine specific amplification of *PIF* elements. Moreover, these elements have conserved TIRs and TSDs (mostly TAA or TTA trinucleotides), have maintained the capacity to code for a transposase as well as the second ORF usually found in *PIF* elements and known as ORF1 or PIFp2 [32–34] (Table 5) and the grapevine database contains a relevant number of ESTs corresponding to *PIF* elements, especially from the *Pifvine-3* and *Pifvine-4* families (Table 5) confirming that these elements are transcribed and potentially active.

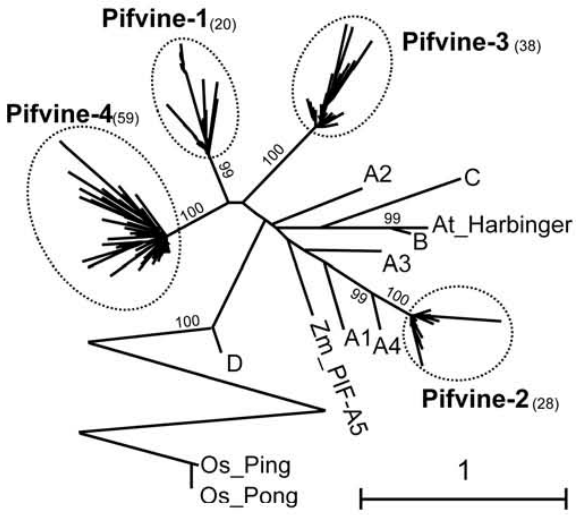

Figure 4. Maximum likelihood tree of the *PIF* superfamily. Bootstrap values higher than 50 are shown. Numbers in brackets show the number of sequences analyzed for each family. Names written in bold are *Vitis* families. Names in plain text are *PIF* elements from other plants (see Materials and Methods for details). The *Ping/Pong* branch is bent to reduce picture size.
doi:10.1371/journal.pone.0003107.g004

Transduplicated cellular gene fragments are present in all superfamilies of *Vitis* class II elements

Transposons can capture host genome sequences and mobilize and amplify them together with their own sequences in a process known as transduplication. Although most of these captured gene fragments seem to be non-functional pseudogenes [31], it has been recently reported that in some cases transduplicated exons could be incorporated into host transcripts by alternative splicing giving rise to new host proteins [35]. Even having lost their coding capacity, transduplicated sequences may undergo transcription and have a regulatory function [31].

MULEs have been shown to frequently capture gene fragments and form Pack-MULEs [36]. MULEs containing transduplicated gene fragments have been reported in *Arabidopsis* [31,37], *Lotus japonicus* [25], melon [38], and rice, were they reach a very high copy number [26,36]. A particular case is the *Arabidopsis KAONASHI*-MULE (*KI*-MULE), a non-TIR MULE found in high copy number that contains a cystein protease domain of 200 amino acids found in ubiquitin-like protein-specific protease (ULP) [31]. In *KI*-MULEs, the ULP protease domain is found in the reverse orientation with respect to the *mudr*A gene. However, examples of ULP-containing MULEs in both direct and reverse orientation have been described also in melon and rice [38]. In addition, the ULP domain in melon can be found in TIR-MULEs and in the distantly related *Jittery*-like MULEs [38]. Our results show that several MULE families identified in grapevine contain sequences with high similarity to ULP genes downstream of the TPase encoding ORF. The ULP coding sequence is found in both orientations in both TIR-MULEs and non-TIR MULEs (Table 4). In addition to *MuDR*-like MULEs, some *Jittery*-like families of grapevine MULEs also contain ULP coding sequences downstream of the transposase ORF (Figure 3). The MULE families containing ULP sequences did not form a monophyletic group (Figures 2A and 2B). In fact, the ULP sequences are found in distantly related elements (MuDR-like and *Jittery*-like), being

Table 5. List of *PIF*-related families of transposons characterized in *Vitis vinifera*.

Family name	Length of complete TE (kb)	N° of TEs having >90% TPase*	N° of TEs with potentially functional ORFs	N° of deleted copies	TIR length in bp	TSD length in bp	N° of EST hits	representative	coordinates
Pifvine-1	5.7	12	4	25	20	3	1	*Harbinger-1_VV*	Repbase
Pifvine-2	7.2	15	5	23	26	3	4	*VHARB-N2_VV*	Repbase
Pifvine-3	6.7–5.8	33	6	25	23	3	10	*Harbinger-3_VV*	Repbase
Pifvine-4	5	33	9	70	35	3	10	*VHARB-N4_VV*	Repbase

* = including the ORF1.
doi:10.1371/journal.pone.0003107.t005

absent in other closely related families, and their presence does not correlate either with the presence or the absence of TIRs, suggesting that ULP transduplication by MULEs is a frequent phenomenon that has occurred independently several times during plant genome evolution. Alternatively, ULP sequences may be frequently lost from MULEs.

In addition to MULEs, *CACTA* elements have also shown to transduplicate cellular genes [39,40], although up to know none has been reported to contain an ULP transduplicated domain. We have found ULP domains in five *CACTA* families (*Cactavine-2*, *Cactavine-3*, *Cactavine-4*, *Cactavine-5* and *Cactavine-13*). We have searched in NCBI for proteins containing the same conserved domain structures as the *CACTA*-ULP found in grapevine and found several proteins from rice that have the Tnp2 and the ULP1 domains. Therefore it appears that *CACTA*-ULPs are common in plants (although perhaps not equally abundant or functional in all genomes since we did not find any similar proteins in *Arabidopsis* or *Medicago* which are genetically closer to *Vitis* than rice is). This also suggests a special "affinity" of the ULP domain to transposons in general.

ULP transduplication is only one example of transduplication. Other genic or non-genic sequences could be "captured" by TEs. For example, in the *Mutators* we have found a family containing intronic and exonic sequences of a putative cellulose synthase gene (Figure 5). In the *CACTAs*, two copies of the *Cactavine-5* family

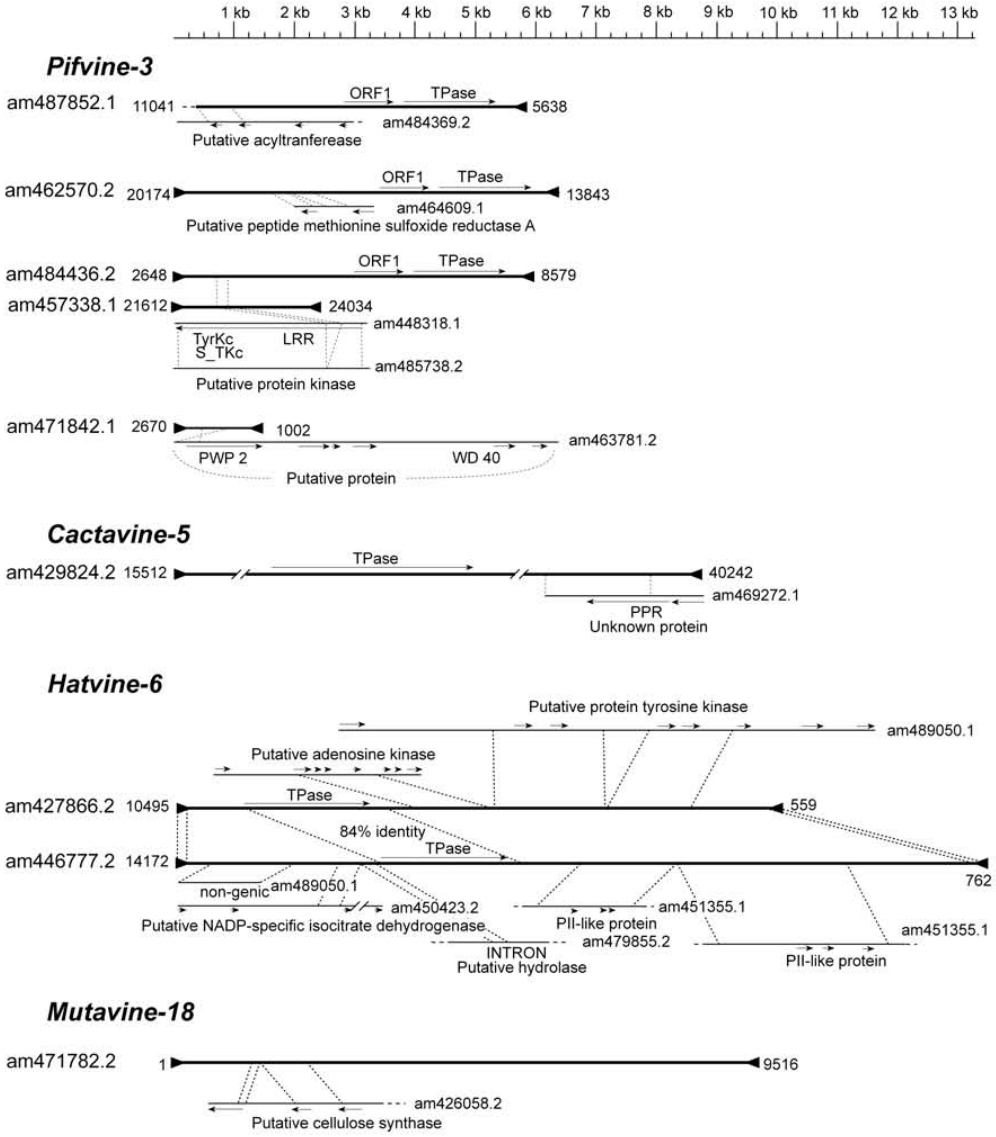

Figure 5. Transduplications of genomic fragments found in different class II elements of *Vitis*. Thick lines represent TEs. Triangles are TIRs. For each source sequence the accession number is given and only for TEs coordinates are given as well. Arrows show the orientation of ORFs. All sequences are draw to the scale.
doi:10.1371/journal.pone.0003107.g005

contain part of the coding sequence and the 3′ untranslated region of a gene encoding for an unknown protein that contains a pentatricopeptide repeat (PPR) domain (Figure 5). This sequence, located downstream of the transposase encoding ORF is found in opposite orientation, and in case of being transcribed from the transposon promoter, would give rise to a transcript antisense to PPR genes with potential regulatory functions.

Although transduplication has only been reported for MULEs and *CACTA* elements in plants, the fact that some of the *PIF* and *hAT* elements here described are unusually long has prompted us to analyze whether these elements contain transduplicated sequences as well. We have analyzed the elements of the *Pifvine-3* family because they very frequently contain a long 5′ region (up to 3.5 kb) that do not correspond to the canonical ORF1 nor transposase coding regions characteristic for these elements. The analysis of these sequences showed that in most cases they share high sequence identity to grapevine genome sequences (including exons and introns) (Figure 5). These transduplications are shared in some cases by multiple copies suggesting that they do not inactivate the transposition of PIF elements. Elements of the *hAT*-family *Hatvine-6* share a similar transposase coding sequence and the TIRs, but the rest of the sequence is often unique, or it is shared by only few elements. Analysis of the variable region of *Hatvine-6* elements revealed that these sequences often share high sequence identity to genic (introns and exons) as well as non-genic grapevine sequences (Figure 5).

Our results show that transduplications are common in grapevine TEs of all superfamilies. We suggest that most plant TEs share this ability as well. Because of their complicated structures and the difficulties to assemble an automated pipeline for their detection, transduplication events are not routinely reported in TE analyses. Thorough analyses, such as the one presented here, are needed to correctly characterize TEs and describe phenomena like the transduplication of cellular sequences.

MULE and *hAT* domesticated transposons

Transposons can lose their ability to transpose and be a source of cellular genes in a process known as domestication. Transposases are specific DNA-binding proteins that catalyze DNA cleavage and strand transfer reactions needed for transposition. Both the DNA binding and the catalytic activity of transposases can be domesticated to give rise to cellular genes [41]. Examples of plant domesticated transposases are the *Arabidopsis* transcription factors *FAR1* and *FHY3*, derived from MULE transposases [42,43] or *DAYSLEEPER*, a gene essential for *Arabidopsis* development which probably encodes a transcription factor derived from a *hAT* transposase [44]. Other domesticated transposons of unknown function are the *MUSTANG* and the *Gary* elements, the former originated from MULE and the later from *hAT* transposons [45,46]. Domesticated transposons are not able to transpose, and for this reason they are in general present as single-copy genes and do not contain TIRs or TSDs.

Five *hAT*-like sequences found in our search are present in single copy and lack TIRs and TSDs: *Vinesleeper-1*, *Vinesleeper-2*, *Hatvine-4*, *Hatvine-5* and *Hatvine-8*. The *Vinesleeper-1* and *Vinesleeper-2* elements are phylogenetically closely related to the *Arabidopsis DAYSLEEPER* (Figure 1) and one of them could be its grapevine orthologue. All 4 ESTs corresponding to *Vinesleeper-1* derive from flower tissues and most of the 11 ESTs corresponding to *Vinesleeper-2* are obtained from different tissues of different developmental stages (Table S1) which suggest a pattern of expression for both genes compatible with a developmentally related function similar to that of *DAYSLEEPER* from *Arabidopsis* [44]. The fact that the grapevine genome contains two potential orthologues for *DAYSLEEPER* suggests that this gene has been duplicated during grapevine evolution and, because of different numbers and origins of corresponding ESTs, the two genes might have diverged to fulfill specialized functions. The other putative domesticated *hAT*-like transposases *Hatvine-4*, *Hatvine-5*, and *Hatvine-8* are not phylogenetically related to *DAYSLEEPER* nor the previously characterized *Gary* element [46]. *Hatvine-8* has a non-functional and partially deleted *TPase* gene which did not allow its alignment and phylogenetical analysis with other members of the *hAT* superfamily, while *Hatvine-4* seems to lack a start codon in its ORF. However, *Hatvine-5* has an intact ORF which matches to transcripts deriving from berry tissue (Table S1) that could be compatible with this element being a domesticated transposase with a function in fruit-related processes.

We have also found MULE-related sequences as candidates for domesticated transposases because of their presence in single copy and lack of TIRs or TSDs (Table 4). These elements belong to the MuDR, *Jittery* and *Hop* families. The MuDR-like elements are phylogenetically closely related to the MUSTANG elements previously described in *Arabidopsis* and sugarcane [45,47] (Figure 3A) and could be the grapevine orthologues of these genes. We have found grapevine ESTs accumulating in different organs and parts of the plant matching to most of these elements (Table 4 and Table S1) which suggests a pattern of expression similar to that of the *Arabidopsis* and sugarcane MUSTANGs [45,47]. Five single copy elements belonging to the *Jittery* family (named *Jithouse*) have been identified (Figure 3B and Table 4) to potentially encode for proteins containing the three domains found in FAR1/FHY3-domesticated transposases (N-terminal C2H2-type zinc-chelating motif of the WRKY-GCM1 family, a central putative core transposase domain and a C-terminal SWIM motif [43]). A recent report has identified 4 out of 5 elements described here as *FRS3*-related *FAR1/FHY3* genes [43]. Although the sequence of *Jithouse-4* was not included in that report, its phylogenetical relationship to the other four elements (Figure 3B) suggests that this is also a *FAR1/FHY3*-related domesticated transposase. Finally, we found one potential domesticated transposase of the *Hop* family, the *Hopvine-2* element present in a single copy and lacks TIRs and TSDs flanking the coding region. The corresponding EST matching to its ORF suggests that *Hopvine-2* be a transposase-related functional gene.

Although the number of ESTs present in grapevine databases is limited for extended expression pattern studies of each putative domesticated element identified, we think the specific nature of these elements could be confirmed. TEs are induced under stress situations, while domesticated transposons lack such a biased expression, most domesticated transposases playing a role in developmentally related processes. 22% of the ESTs corresponding to the putative domesticated transposases here described belong to EST collections obtained from stressed material, which is almost exactly the percentage of the stress-related EST collections in the total grapevine EST databases (23%). Contrastingly, 77% of the ESTs corresponding to potentially mobile transposons are obtained from stressed material which is significantly more than expected (χ^2 test, pvalue$<$0.0001). This difference in expression confirms the classification as true transposons and domesticated transposases made here based on molecular characteristics.

Insertion polymorphisms of grapevine cut-and-paste transposons revealed by PCR

The results presented here show that a high number of grapevine transposons have maintained the capacity to encode a transposase and are expressed under particular situations,

suggesting that they may have retained the capacity to transpose. In order to get more information on the possible mobility of these elements, we looked for insertion polymorphisms of eleven of these elements among seven grapevine cultivars. We have also included in this analysis four putative domesticated elements which are supposed to have lost their ability to transpose. The presence of a given element at a particular location in the genome was revealed by a PCR amplification using a primer complementary to the internal region of the TE and a primer designed in the flanking region. To check for the absence of a given element at a particular location we performed PCR amplifications with two primers complementary to the regions flanking the element at both sides (see Materials and Methods for details). Some randomly chosen bands were sequenced to confirm the nature of the amplification products.

None of the four putative domesticated transposases analyzed showed insertion polymorphisms (Figure 6, bottom panel). Taking into account the high heterozygosity of grapevine this result suggests that domesticated transposons fulfill important cellular roles and have been under strong selective pressure for their maintenance. On the contrary, all but one (Hatvine-7.1) transposon insertions analyzed are polymorphic (8 examples are shown in Figure 6, top and middle panels). This could suggest that most transposon insertions are not under strong selective pressure and are randomly distributed among cultivars. Alternatively, this result may also indicate that some of these insertions are recent and have not had time to become fixed. In particular, Pifvine-2 insertions could be relatively recent (possibly after the domestication of grapevine), as only two out of seven cultivars contain the insertion at this particular locus (Figure 6). In some cases we obtained multiple bands, or products with unexpected sizes. The sequence of the unexpectedly small bands of the *Pifvine-2* empty sites (for samples 4 and 6) and the unusually bigger band of the *Pifvine-3* empty site (sample 5) revealed sequence polymorphisms

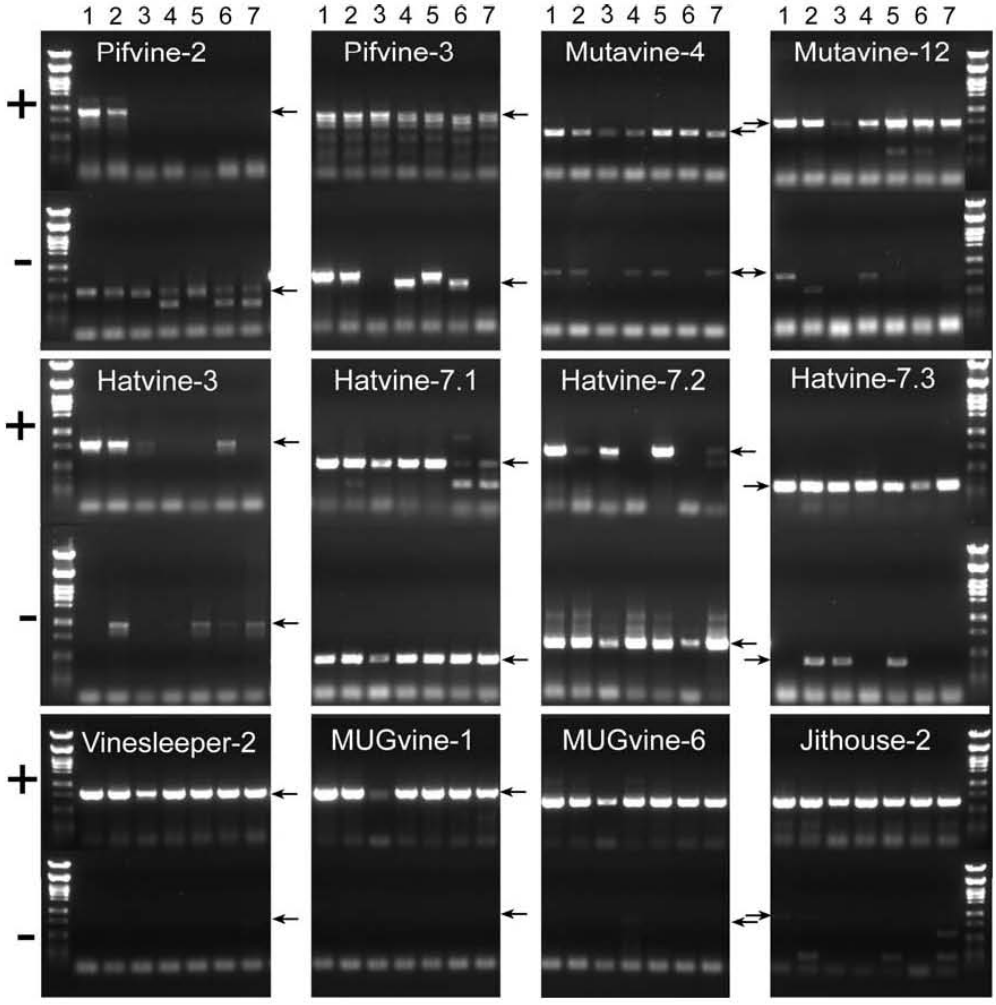

Figure 6. Examples of the insertion polymorphism of different TEs and domesticated transposases from grapevine. The culivars analyzed are Pinot Noir (1), Riesling (2), Chardonnay (3), Cabernet Sauvignon (4), cabernet Mitos (5), Cabernet Cortis(6) and cabernet Carbon (7). "+" indicate the insertion at a given locus, while "−" indicate an empty site. Arrows indicate the expected size of the band. Numbers are grapevine cultivars (in the same order as given in Table S2).
doi:10.1371/journal.pone.0003107.g006

unrelated to the transposition of the elements here reported. In the case of *Pifvine-2* we found a 154 bp-long deletion present 216 bp downstream of the target site, while in the case of *Pifvine-3* there is an insertion of a putative SINE element (155 bp-long with 13 bp-long TSDs) 22 bp after the target site.

This results thus show that a high proportion of grapevine "cut-and-paste" transposons have recently transposed during grapevine evolution, accompanying its domestication and breeding processare polymorphic and contribute to the high variability of grapevine genome.

Conclusions

We have performed a detailed analysis of the "cut-and-paste" transposons of *Vitis vinifera* L, and found that this genome contains elements belonging to four of the five superfamilies of elements described in plants, *hAT*, *CACTA*, *Mutator* and *PIF*. *hAT* and *Mutator* superfamilies are the most prevalent in grapevine, while *CACTA* is probably the superfamily that has had the less activity in the recent grapevine genome evolution. The presence of TSDs, intact ORFs and high number of corresponding ESTs, as well as the high frequency of insertion polymorphisms among different grapevine cultivars show that these elements have transposed recently during grapevine evolution and suggests that some of them may have retained the capacity to transpose. On the contrary, the genome of grapevine also contains an important number of domesticated transposases belonging to different superfamilies that have lost the ability to transpose and probably fulfill cellular functions. Additionally, we found that transduplication of gene fragments is not restricted only to MULEs and CACTAs but can occur in other superfamilies as well. Our results show that, as in most complex genomes, TEs have made an important contribution to grapevine genome evolution and variation today.

Materials and Methods

Transposon mining

We performed our analyses using the whole genome shotgun sequences of the grapevine genome made available at NCBI by Velasco et al. in January 2007 [13]. Sequences from Jaillon et al. [12] were made available at NCBI after we had started with our analyses and were used as confirmation references. As a first approach to characterize grapevine class II "copy-and-paste" transposons we used a homology-based strategy to look for sequences with similarities with known transposases. We retrieved protein sequences of plants from NCBI (in May 2007) using keywords as "transposase" or class II superfamily names like "Mutator", "MUDRA", "CACTA", "hAT" etc. We grouped the retrieved transposase sequences into belonging superfamilies and performed a blastx search [48] with the grapevine genome shotguns as queries. We considered all shotguns having an e-value lower than 1×10^{-50} for their best TPase hit. These shotguns were manually checked and the putative TPase was analyzed. TPase genes were characterized by blastx of the shotgun of interest to the whole NCBI protein database. In this way, similarities with non-annotated proteins could be determined as well. As both [12] and [13] performed computational gene predictions, the NCBI contains a significant number of predicted (but not annotated) *Vitis* proteins which were useful to precisely determine the borders of putative TPase for each TE family analyzed. The TPase regions with several kb of flanking sequence were blasted against the whole *Vitis* shotgun database to determine the full length or the borders of the element. TIRs were manually looked for, or by using the FastPCR software (Kalendar 2006, www.biocenter.helsinki.fi/bi/programs/fastpcr.htm). By blasting the putative full length element to the *Vitis* whole genome shotgun database we could also find non-autonomous or deleted elements of the same family which have lost the TPase gene. To quantify all sequences belonging to the same family we used a full length element as query and considered all fragments with at least 80% identity and having at least 20% of the query length. We used the rule of >80% sequence similarity to group elements into the same family.

Phylogeny of the TEs

Each TE superfamily was phylogenetically analyzed to determine the number and relationships of the families and to compare them to some known elements from other plants. We aligned amino-acid sequences of conserved TPase regions using ClustalW algorithm [49] implemented in the BioEdit software [50]. PHYML software [51] was used to build phylogenies using maximum likelihood with the JTT model of evolution, four substitution rate categories, fixed proportion of invariable sites and non parametric bootstrap analysis of 100 replicates.

For the *hAT* superfamily we used a 39 aa-long region as in [52]. For comparison with -*Vitis* elements -we included the following *hAT* TEs in the phylogenetical tree: *AC9* (accession No X05424), *Bg* (accession No X56877), *Tag1* (accession No AAC25101), *Tam3* (accession No X55078). We also included the domesticated TPases *DAYSLEEPER* [44] and *r-gary1* [46]. The multiple alignments are given in Dataset S5.

For the *CACTA* superfamily we used amino-acid fragments homologous to the *En-1* TPase (accession No AAA66266), between positions 287 and 435. For comparison with *Vitis* elements we included the following elements in the phylogenetical tree: *PSL* (accession number AF009516), *ATENSPM2* [54], *Doppia4* (accession No AF187822), *En1* (accession No AAA66266), *TNP2* (accession No CAA40555.1) and *OSHOOTER* [55]. The multiple alignments are given in Dataset S6.

For the *Mutator* superfamily we used amino-acid fragments homologous to MURA between positions 468 and 640 as in Saccaro et al., [47] . For comparison with *Vitis* elements we included MURA, TE165, OsMUG1, SCMUG263, SCMUG228, AtMUG1, AtMUG05a, AtMUG05b, AtMUG03b, AtMUG03c [47] and MuDR2_OS [53]. The multiple alignments are given in Dataset S7. Comparison between *MuDR*-like and *Jittery/Hop*-like elements was possible only by comparing the amino-acid fragments homologous to *Jittery* TPase (accession No AAF66982) between positions 217 and 343 and *Hop* (accession No AAP31248.1) between positions 203 and 331. The only *MuDR*-like elements form *Vitis* that could be aligned with *Jittery* and *Hop* were *Mutavine-1*, *12* and *14* as well as *MUGvine-5*. The multiple alignments are given in Dataset S8.

For the *PIF* superfamily we used amino-acid fragments as described in Figure 1 in Zhang et al. [32] . For comparison with *Vitis* elements we included Os_Pong and Os_Ping and representatives from each *PIF* cluster from the Figure 3 in Zhang et al. [32]: HvBF628721 for cluster A1, ShAY362818 for cluster A2, AtAC007123 for cluster A3, LjAP004528 for cluster A4, Zm_PIF for cluster A5, BoBH561775 for cluster B, BoBH485472 for cluster C and ZmAF072725 for cluster D. In addition we included *Harbinger* [54]. The multiple alignments are given in Dataset S9.

All trees were visualized using *MEGA* version 3.1. [56]

Submission to Repbase Reports

For some families having true full length individual copies (with TSDs and/or TIRs and the coding region) consensus sequences were created and submitted to Repbase Reports (http://www.girinst.org/repbase/). Names were changed according to the new Repbase nomenclature (Tables 1–5).

Plant material

A list of samples and their source is given in Table S2. DNA from all samples was extracted using E.Z.N.A. SP Plant DNA Mini Kit (Omega Bio-tek).

PCR analysis

Primers were designed using FastPCR software (Kalendar 2006, www.biocenter.helsinki.fi/bi/programs/fastpcr.htm). Each primer was blasted against the whole *Vitis* genomic database to check for specificity. The list of primers is given in Table S3. PCRs were done in 20 µl reaction volumes using approximately 30 ng of template DNA, 0.5 µl of each primer (10 pmol/µl), and TaKaRa Ex Taq in the following conditions: 94 °C·2 min^{-1}+40×(94 °C·25 s^{-1}, 59 °C·45 s^{-1}, 72 °C·1 min^{-1})+72 °C·5 min^{-1}. PCR products were run in 1.2% agarose gels with EtBr in a 1× TAE buffer and visualized under UV light.

Supporting Information

Table S1 Detailed information of TEs and ESTs from grapevine.
Found at: doi:10.1371/journal.pone.0003107.s001 (0.15 MB DOC)

Table S2 List of samples used for the PCR analysis.
Found at: doi:10.1371/journal.pone.0003107.s002 (0.03 MB DOC)

Table S3 The list of primers used for insertion polymorphism analysis.
Found at: doi:10.1371/journal.pone.0003107.s003 (0.04 MB DOC)

Dataset S1 Supporting information on the hAT superfamily
Found at: doi:10.1371/journal.pone.0003107.s004 (0.50 MB XLS)

Dataset S2 Supporting information on the CACTA superfamily
Found at: doi:10.1371/journal.pone.0003107.s005 (0.12 MB XLS)

Dataset S3 Supporting information on the Mutator superfamily
Found at: doi:10.1371/journal.pone.0003107.s006 (0.43 MB XLS)

Dataset S4 Supporting information on the PIF superfamily
Found at: doi:10.1371/journal.pone.0003107.s007 (0.22 MB XLS)

Dataset S5 Multiple alignments used for the phylogenetical analysis of hAT elements.
Found at: doi:10.1371/journal.pone.0003107.s008 (0.05 MB DOC)

Dataset S6 Multiple alignments used for the phylogenetical analysis of CACTA elements.
Found at: doi:10.1371/journal.pone.0003107.s009 (0.03 MB DOC)

Dataset S7 Multiple alignments used for the phylogenetical analysis of Mutator elements.
Found at: doi:10.1371/journal.pone.0003107.s010 (0.10 MB DOC)

Dataset S8 Multiple alignments used for the phylogenetical analysis of Jittery-like and Hop-like elements.
Found at: doi:10.1371/journal.pone.0003107.s011 (0.01 MB DOC)

Dataset S9 Multiple alignments used for the phylogenetical analysis of PIF elements.
Found at: doi:10.1371/journal.pone.0003107.s012 (0.02 MB DOC)

Acknowledgments

For providing the plant material we thank Ernst Rühl (Institute of Grapevine Breeding Geisenheim, Germany), Bernd Hill (LVWO Weinsberg, Germany), and Reinhard Antes (Nursery Antes, Heppenheim, Germany).

Author Contributions

Conceived and designed the experiments: AB JMC. Performed the experiments: AB. Analyzed the data: AB JMC. Contributed reagents/materials/analysis tools: AF. Wrote the paper: AB AF JMC.

References

1. Levadoux L (1956) Les populations sauvages et cultivées de Vitis vinifera L. Ann Amélior Plant 6: 59 117.
2. Arroyo-Garcia R, Ruiz-Garcia L, Bolling L, Ocete R, Lopez MA, et al. (2006) Multiple origins of cultivated grapevine (Vitis vinifera L. ssp. sativa) based on chloroplast DNA polymorphisms. Molecular Ecology 15: 3707 3714.
3. This P, Lacombe T, Thomas MR (2006) Historical origins and genetic diversity of wine grapes. Trends in Genetics 22: 511 519.
4. Forneck A (2005) Plant Breeding: Clonality - A concept for stability and variability during vegetative propagation. In: Esser ULK, Beyschlag W, Murata J, eds. Progress in Botany. Heidelberg, Germany: Springer Berlin. pp 165 183.
5. Franks T, Botta R, Thomas MR, Franks J (2002) Chimerism in grapevines: implications for cultivar identity, ancestry and genetic improvement. TAG Theoretical and Applied Genetics 104: 192 199.
6. Le Q, Melayah D, Bonnivard E, Petit M, Grandbastien M (2007) Distribution dynamics of the Tnt1 retrotransposon in tobacco. Molecular Genetics and Genomics 278: 1617 4615.
7. Kobayashi S, Goto-Yamamoto N, Hirochika H (2004) Retrotransposon-Induced Mutations in Grape Skin Color. Science 304: 982-.
8. Lijavetzky D, Ruiz-García L, Cabezas J, De Andrés M, Bravo G, et al. (2006) Molecular genetics of berry colour variation in table grape. Molecular Genetics and Genomics 276: 427 435.
9. Walker A, Lee E, Bogs J, McDavid D, Thomas M, et al. (2007) White grapes arose through the mutation of two similar and adjacent regulatory genes. The Plant Journal 49: 772 785.
10. Pelsy F, Merdinoglu D (2002) Complete sequence of Tvv1, a family of Ty1 copia-like retrotransposons of Vitis vinifera L., reconstituted by chromosome walking. TAG Theoretical and Applied Genetics 105: 614 621.
11. 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.
12. Jaillon O, Aury JM, Noel B, Policriti A, Clepet C, et al. (2007) The grapevine genome sequence suggests ancestral hexaploidization in major angiosperm phyla. Nature 449: 463 467.
13. Velasco R, Zharkikh A, Troggio M, Cartwright D, Cestaro A, et al. (2007) A High Quality Draft Consensus Sequence of the Genome of a Heterozygous Grapevine Variety. PLoS ONE 2: e1326.
14. Feschotte C, Pritham EJ (2007) DNA Transposons and the Evolution of Eukaryotic Genomes. Annual Review of Genetics 41: 331 368.
15. Guynet C, Hickman AB, Barabas O, Dyda F, Chandler M, et al. (2008) In Vitro Reconstitution of a Single-Stranded Transposition Mechanism of IS608. Molecular Cell 29: 302 312.
16. Barabas O, Ronning DR, Guynet C, Hickman AB, Ton-Hoang B, et al. (2008) Mechanism of IS200/IS605 Family DNA Transposases: Activation and Transposon-Directed Target Site Selection. Cell 132: 208 220.
17. Wicker T, Sabot F, Hua-Van A, Bennetzen JL, Capy P, et al. (2007) A unified classification system for eukaryotic transposable elements. Nat Rev Genet 8: 973 982.
18. Feschotte C, Osterlund MT, Peeler R, Wessler SR (2005) DNA-binding specificity of rice mariner-like transposases and interactions with Stowaway MITEs. Nucl Acids Res 33: 2153 2165.
19. Rubin E, Lithwick G, Levy AA (2001) Structure and Evolution of the hAT Transposon Superfamily. Genetics 158: 949 957.
20. Zhang X, Wessler SR (2004) Genome-wide comparative analysis of the transposable elements in the related species Arabidopsis thaliana and Brassica oleracea. Proceedings of the National Academy of Sciences 101: 5589 5594.

21. Wicker T, Guyot R, Yahiaoui N, Keller B (2003) CACTA Transposons in Triticeae. A Diverse Family of High-Copy Repetitive Elements. Plant Physiol 132: 52 63.
22. Miura A, Kato M, Watanabe K, Kawabe A, Kotani H, et al. (2004) Genomic localization of endogenous mobile CACTA family transposons in natural variants of Arabidopsis thaliana. Molecular Genetics and Genomics 270: 524 532.
23. Robertson D (1978) Characterization of a mutator system in maize. Mutant Research 51: 21 28.
24. Lisch D (2002) Mutator transposons. Trends in Plant Science 7: 498 504.
25. Holligan D, Zhang X, Jiang N, Pritham EJ, Wessler SR (2006) The Transposable Element Landscape of the Model Legume Lotus japonicus. Genetics 174: 2215 2228.
26. Juretic N, Hoen DR, Huynh ML, Harrison PM, Bureau TE (2005) The evolutionary fate of MULE-mediated duplications of host gene fragments in rice. Genome Res 15: 1292 1297.
27. Turcotte K, Srinivasan S, Bureau T (2001) Survey of transposable elements from rice genomic sequences. The Plant Journal 25: 169 179.
28. Xu Z, Yan X, Maurais S, Fu H, O'Brien DG, et al. (2004) Jittery, a Mutator Distant Relative with a Paradoxical Mobile Behavior: Excision without Reinsertion. Plant Cell 16: 1105 1114.
29. Chalvet F, Grimaldi C, Kaper F, Langin T, Daboussi M-J (2003) Hop, an Active Mutator-like Element in the Genome of the Fungus Fusarium oxysporum. Mol Biol Evol 20: 1362 1375.
30. Le QH, Wright S, Yu Z, Bureau T (2000) Transposon diversity in Arabidopsis thaliana. Proceedings of the National Academy of Sciences 97: 7376 7381.
31. Hoen DR, Park KC, Elrouby N, Yu Z, Mohabir N, et al. (2006) Transposon-Mediated Expansion and Diversification of a Family of ULP-like Genes. Mol Biol Evol 23: 1254 1268.
32. Zhang X, Jiang N, Feschotte C, Wessler SR (2004) PIF- and Pong-Like Transposable Elements: Distribution, Evolution and Relationship With Tourist-Like Miniature Inverted-Repeat Transposable Elements. Genetics 166: 971 986.
33. Casola C, Lawing AM, Betran E, Feschotte C (2007) PIF-like Transposons are Common in Drosophila and Have Been Repeatedly Domesticated to Generate New Host Genes. Mol Biol Evol 24: 1872 1888.
34. Kapitonov VV, Jurka J (2004) Harbinger Transposons and an Ancient HARBI1 Gene Derived from a Transposase. DNA and Cell Biology 23: 311 324.
35. Zabala G, Vodkin L (2007) Novel exon combinations generated by alternative splicing of gene fragments mobilized by a CACTA transposon in Glycine max. BMC Plant Biology 7: 38.
36. Jiang N, Bao Z, Zhang X, Eddy SR, Wessler SR (2004) Pack-MULE transposable elements mediate gene evolution in plants. Nature 431: 569 573.
37. Yu Z, Wright SI, Bureau TE (2000) Mutator-like Elements in Arabidopsis thaliana: Structure, Diversity and Evolution. Genetics 156: 2019 2031.
38. van Leeuwen H, Monfort A, Puigdomenech P (2007) Mutator -like elements identified in melon, Arabidopsis and rice contain ULP1 protease domains. Molecular Genetics and Genomics 277: 357 364.
39. Kawasaki S, Nitasaka E (2004) Characterization of Tpn1 Family in the Japanese Morning Glory: En/Spm-related Transposable Elements Capturing Host Genes. Plant Cell Physiol 45: 933 944.
40. Zabala G, Vodkin LO (2005) The wp Mutation of Glycine max Carries a Gene-Fragment-Rich Transposon of the CACTA Superfamily. Plant Cell 17: 2619 2632.
41. Volff J-N (2006) Turning junk into gold: domestication of transposable elements and the creation of new genes in eukaryotes. BioEssays 28: 913 922.
42. Hudson ME, Lisch DR, Quail PH (2003) The FHY3 and FAR1 genes encode transposase-related proteins involved in regulation of gene expression by the phytochrome A-signaling pathway. The Plant Journal 34: 453 471.
43. Lin R, Ding L, Casola C, Ripoll DR, Feschotte C, et al. (2007) Transposase-Derived Transcription Factors Regulate Light Signaling in Arabidopsis. Science 318: 1302 1305.
44. Bundock P, Hooykaas P (2005) An Arabidopsis hAT-like transposase is essential for plant development. Nature 436: 282 284.
45. Cowan RK, Hoen DR, Schoen DJ, Bureau TE (2005) MUSTANG Is a Novel Family of Domesticated Transposase Genes Found in Diverse Angiosperms. Mol Biol Evol 22: 2084 2089.
46. Muehlbauer G, Bhau B, Syed N, Heinen S, Cho S, et al. (2006) A hAT superfamily transposase recruited by the cereal grass genome. Molecular Genetics and Genomics 275: 553 563.
47. Saccaro JNL, Van Sluys M-A, de Mello Varani A, Rossi M (2007) MudrA-like sequences from rice and sugarcane cluster as two bona fide transposon clades and two domesticated transposases. Gene 392: 117 125.
48. Altschul SF GW, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. J Mol Biol 215: 403 410.
49. Thompson JD, Higgins DG, Gibson TJ (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res 22: 4673 4680.
50. Hall TA (1999) BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. Nucl Acids Symp Ser 41: 95 98.
51. Guindon S, Lethiec F, Duroux P, Gascuel O (2005) PHYML Online a web server for fast maximum likelihood-based phylogenetic inference. Nucl Acids Res 33: W557 559.
52. Kempken F, Windhofer F (2001) The hAT family: a versatile transposon group common to plants, fungi, animals, and man. Chromosoma 110: 1 9.
53. Jurka J (2005) MuDR2_OS: MuDR-type DNA transposon from rice. Repbase Reports 5: 201 201.
54. Kapitonov V, Jurka J (1999) Molecular paleontology of transposable elements from Arabidopsis thaliana. Genetica 107: 27 37.
55. Jurka J (2005) OSHOOTER: EnSpm-type DNA transposon from rice. Repbase Reports 5: 205.
56. Kumar S, Tamura K, Nei M (2004) MEGA3: Integrated software for Molecular Evolutionary Genetics Analysis and sequence alignment. Briefings in Bioinformatics 5: 150 163.

ARTICLE 6

Recent amplification and impact of MITEs on the genome of grapevine (*Vitis vinifera* L.)

Andrej Benjak, Stéphanie Boué, Astrid Forneck, Josep M. Casacuberta

This work was done in collaboration with Dr. Josep M. Casacuberta at the CRAG, Barcelona, Spain. The article presented here is a draft manuscript which will be soon submitted for publication in a scientific journal. The manuscript is subject to correction and modifications prior to publication and it is shown here only for educational purposes.

Recent amplification and impact of MITEs on the genome of grapevine (Vitis vinifera L.)

Andrej Benjak[1, 2], Stéphanie Boué[3], Astrid Forneck[2], Josep M. Casacuberta[1]

[1] Department of Plant Molecular Genetics, Center for Research in Agricultural Genomics (CRAG, CSIC-IRTA-UAB), Barcelona, Spain.

[2] Institute of Horticulture and Viticulture, University of Natural Resources and Applied Life Sciences, Vienna, Austria.

[3] Center of Regenerative Medicine in Barcelona (CMR[B]), Barcelona, Spain.

Introduction

Miniature inverted-repeat transposable elements (MITEs) are a particular type of defective class II transposons. They share some features with non-autonomous class II transposons: they are characterized by their terminal inverted repeat (TIR) structure, short direct repeats formed by target site duplication (TSD) and absence of gene-coding capacity. Most MITE families share extensive sequence similarities with class II transposons from which they are supposed to derive by internal deletion (Feschotte and Mouches 2000, Jiang *et al.* 2004, Moreno-Vázquez *et al.* 2005, Yang and Hall 2003, Zhang *et al.* 2004). On the other hand, MITEs are distinguished from other non-autonomous class II transposons by their high copy number, the high uniformity of their copies and in some cases with the potential to form single strand secondary structures. These features allow them to influence the genome in different ways, from insertional mutagenesis to gene regulation.

MITE mobility causes insertion polymorphism and increases the allelic diversity of the genome (Feschotte and Pritham 2007). Moreover, MITEs are often found close or within genes. Like for other short transposable elements (TEs), insertion of MITEs in introns or close to genes is less deleterious compared to long TEs. On the other hand there are extreme examples of MITE families that have strong preference for insertion into genes or regulatory regions. For example, especially biased for such insertions are the *Glider* and

Vision MITE families from *Xenopus laevis* where almost all copies (there are 20,000 copies of *Glider*) are inserted into regulatory regions (Lepetit *et al.* 2000). MITEs that are inserted into or close to genes can affect these genes by providing for new splicing sites, transcription start sites, new exons, and poly(A) sites (Kuang *et al.* 2008, Ohmori *et al.* 2008, Santiago *et al.* 2002). Additionally, MITEs can give rise to siRNA genes and regulate genes that are not necessary in their proximity (Kuang, et al. 2008, Piriyapongsa and Jordan 2007, Piriyapongsa and Jordan 2008). These phenomena show that MITEs have strong repercussions in the evolution of genes and the plasticity of the genomes.

Grapevine is a widely cultivated crop and a high number of different varieties have been selected since its domestication in the Neolithic period. Although sexual crossing has been a major driver of grapevine evolution, its vegetative propagation enhanced the impact of somatic mutations and has been important for grapevine diversity. Transposable elements are known to be major contributors to genome variability and, in particular, to somatic mutations. Thus, transposable elements have probably played a major role in grapevine domestication and breeding. We recently and for the first time described 51 families of class II transposons in grapevine and 15 families of putative domesticated transposons (Benjak *et al.* 2008). We showed that class II TEs may have contributed in shaping the grapevine genome through insertion polymorphism, gene transduplication and TE domestication. In this work we analyze the MITE subfamilies which directly evolved as deletion derivates from the TE families described before in grapevine (Benjak, et al. 2008) and provide evidences for their major role in shaping the grapevine genome.

Results and discussion

Grapevine contains MITEs related to different superfamilies of class II transposons

Although the first MITEs to be described in plants were related to elements of the *PIF*/*Pong* and *Mariner* families (Feschotte and Mouches 2000, Feschotte *et al.* 2003, Zhang, et al. 2004) MITEs related to most class II families of TEs have been described in plants later on (Kuang, et al. 2008, Saito *et al.* 2005, Yang and Hall 2003). We have recently described the "cut-and-paste" transposon landscape of the grapevine genome, and found that it contains

representatives of 4 of the 5 superfamilies present in plants (Benjak, et al. 2008). Here we present an analysis of the MITEs directly associated to the described grapevine transposons. We searched for elements sharing TIRs and subterminal sequences with those elements, devoid of transposase coding capacity and being present as a high number of copies highly homogeneous in size and sequence. This was done by visual inspection of blast results from searches with representatives of the each of the major families of class II transposons previously described in *Vitis* (Benjak, et al. 2008). We have not found any potential MITEs related to grapevine MULEs. Although we found defective MULEs, each defective element was not present in multiple copies. On the contrary, we have found potential MITEs related to the other three transposon superfamilies present in grapevine, the *CACTA*, *hAT* and *PIF* superfamilies, and we named these putative MITEs according to the previously given family names (m-"TE family name".MITE subfamily number). These elements, related to 8 families of transposons, are highly homogeneous in size and sequence, which suggest that they have been amplified from a single or few founder elements, as it is usually the case for MITEs (Casacuberta and Santiago 2003, Deragon *et al.* 2008, Feschotte *et al.* 2002). Seven of these subfamilies (*mCactavine-4.1*, *mHatvine-2.n*, *mHatvine-3.1*, *mHatvine-10.1*, *mPifvine-1.1*, *mPifvine-2.1* and *mPifvine-4.1*) are composed by a moderate copy number of relatively long elements, and only *mPifvine-3.1* elements are present at high copy number and are of a size similar to the typical MITE families described in plants (Table 1).

Table 1: MITEs found in grapevine and their properties

MITE name	Average length in base pairs (standard deviation)	Approx. copy number	Total coverage in kb	TIR length	TSDs length	Average identity	ESTs matching to MITEs	Representative	Coordinates
mPifvine-1.1	715 (δ=27.2)	51	37.2	20	3	0.88	1	AM485510.1	38610-37877
*mPifvine-2.n**	~1 kb	54	91.7	26	3	0.77	2	AM452748.2	1881-2798
mPifvine-3.1	274 (δ=10.8)	1298	355.8	18	3	0.83	~50	AM468072.2	15068-15345
mPifvine-4.1	1243 (δ=90.2)	76	94.5	11	3	0.81	1	AM450168.2	3569-4849
mHatvine-2.1	769 (δ=14.4)	22	16.9	23	8	0.92	2	AM436283.2	16403-17173
mHatvine-3.1	740 (δ=77.7)	30	22.2	16	8	0.71	3	AM458859.2	11690-12492
mHatvine-10.1	1274 (δ=23.7)	20	22.9	11	8	0.89	0	AM432725.2	460-1725
mCactavine-4.1	3243 (δ=23.8)	30	97.3	6	3	0.84	31	AM457287.1	34741-37976

*We have described 13 subfamilies of *mPifvine-2* MITEs (see Table 2).

Pack-MITEs: MITEs transduplicating gene sequences

Although all the MITE families here described were found because they show extensive sequence similarity with the related TE families, some of them (*mCactavine-4.1*, *mHatvine-10.1*, *mPifvine-2.n* and *Pifvine-4.1*) also contain sequences unrelated to their corresponding long elements. In the case of *mCactavine-4.1,* the internal sequence, which is 900 bp-long and is highly conserved in all copies, does not have similarity to any other sequence (not shown). For the other MITE subfamilies the internal sequences are highly similar to grapevine genomic sequences. This suggests that MITEs can capture, mobilize and amplify host genomic sequences as typical DNA transposons do, in a process that has been named transduplication (Juretic *et al.* 2005). We have thus analyzed these possible examples of transduplication with some detail. *mHatvine-10.1* elements share TIRs and sub-terminal regions with *Hatvine-10* TEs, but contain a central 583 bp-long region not related to *Hatvine-10* but to a cellular non-genic region (82% identity, Figure 1). This region is highly conserved in all 20 *mHatvine-10.1* elements (average identity 88.5%) and suggests that an ancestral *mHatvine-10.1* transduplicated a genomic region and that the composite element was later on amplified. A similar scenario could also explain the structure of *mPifvine-2* elements which also contain a central region unrelated to *Pifvine-2* transposons. However, this central sequence is not the same in all *mPifvine-2* copies, as each internal sequence is shared only by few elements. We grouped these elements in 13 subfamilies, each of which has a different transduplicated sequence. In all the cases the internal sequence shows high sequence similarity to a grapevine genomic sequence (Table 2) suggesting that all have been transduplicated by *mPifvine-2* elements (Figure 1 and Table 2). Most transduplicated sequences are coding sequences corresponding to expressed grapevine genes, and fragments from different genes can be present in a single *mPifvine-2* element (Figure 1 and Table 2). Interestingly, in the cases when a subfamily has transduplicated several gene fragments, elements with a different number of such fragments can be distinguished, suggesting that a subset of elements containing transduplicated gene fragments have undergone additional rounds of transduplication and amplification (Figure 1).

In addition to the abovementioned examples of transduplication, we have characterized another example of transduplication that has particular characteristics. *mPifvine-4.1* MITEs, present in more than 70 copies which are around 1,200 bp-long (Table 1), have a 780 bp-long central region which is not found in the full length *Pifvine-4* elements. Differently to most transduplicated sequences, this central region does not correspond to a single copy sequence found elsewhere in the genome, but to a repetitive sequence, present in more than 180 copies in grapevine that seems to be highly expressed as it matches to more than 100 ESTs deposited in the grapevine EST databases (not shown). This repetitive sequence is 1.5 Kb long and is flanked by direct repeats of 7 bp that could represent TSD generated upon insertion. All these characteristics suggest that this sequence, which we have named *Mila* (its reference is given in Figure 1), is a novel mobile element as it seems to be different in sequence and structure from other known TEs. Because the sequence included within *mPifvine-4* elements is only part of *Mila* (which also contains a partial tandem duplication of a central motif), we suggest that *Mila* sequences have been transduplicated and further amplified by *mPifvine-4.1* elements.

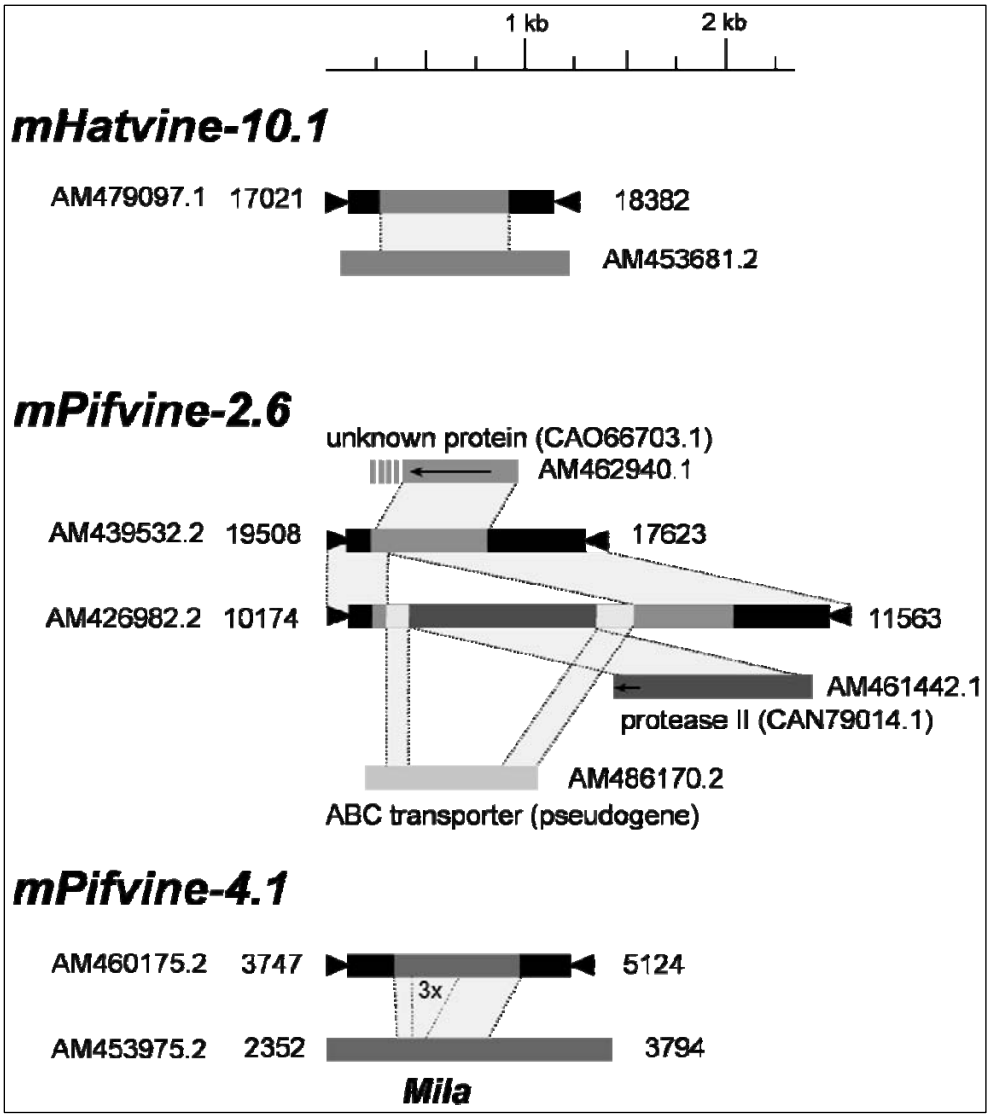

Figure 1: Transduplication of genomic sequences by MITEs. Triangles represent TIRs. Arrows represent exons. Accession numbers for all sequences are given as well as the coordinates for the transposable elements.

Table 2: Information on transduplicated sequences in some *mPifvine-2* subfamilies.

mPifvine-2 subfamily	Copy number	Accession nº for the representative	Coordinates	Trans-duplication in bp	Genomic source of transduplication	Identity of the original sequence to its transduplication	Predicted protein	Tentative protein annotation	Trans-duplicated exons
1	6	AM452748.2	1881-2798	500	AM462940.1	88%	-	Thaumatin-like	yes
2	4	AM442314.2	17623-19508	1300	AM449452.1	94%	CAO21921.1	BAH-AAA-containing protein	yes
3	4	AM477430.2	3327-5005	800	AM441647.2	93%	CAN62993.1	Glycosyl transferase	yes
4	6	AM426982.2	5977-8628	700	AM486170.2	91%	-	ABC transporter (pseudogene)	yes
				1000	AM461442.1	86%	-	Protease II	yes
				95	AM475940.1	91%	CAO66703.1	Unknown protein	yes
5	1	AM447383.2	3750-4593	400	CU459360.1	83%	CAO66902	DUF1296-containing protein	yes
6	2	AM467559.2	7724-9445	700	AM437259.2	91%	CAN82620.1	Glycosyl hydrolase	yes
7	6	AM436343.2	7645-9489	~1000	?	-	similar to CAO46980.1	Unknown protein	?
8	6	AM449479.2	9036-10309	80	?	-	similar to CAO40038.1	Sec15-containing protein	yes
				100	?	-	similar to CAO49645.1	Unknown protein	yes
				600	AM463124.2	91%	CAO39675.1	RING finger	yes
9	12	AM445491.2	6050-7622	830	AM452971.1	82%	-	-	no
10	4	AM429887.2	6356-7627	630	AM425582.2	92%	-	-	no
11	2	AM432699.2	8138-9741	700	AM458836.2	92%	CAN72319.1	Serine protease	yes
12	4	AM431974.2	47527-48887	860	AM471293.1	92%	CAO46017.1	Sulfate transporter like protein	yes
13	2	AM433436.2	1665-2724	545	AM450890.2	97%	CAO42680.1	Pectinesterase	yes

Recent transposition and amplification of *mPifvine-3.1* MITEs in *Vitis* species

Although we have described putative MITEs belonging to 8 different TE families, only the *Pifvine-3* related MITE subfamily (named *mPifvine-3.1*) is present in high copy number and has all the hallmarks of these elements. A phylogenetic analysis of *mPifvine-3.1* elements did not allow grouping them into distinct clusters (not shown) suggesting that *mPifvine-3.1* elements have not amplified from different bursts of amplification of distinct founder elements. The fact that the some 1,200 *mPifvine-3.1* are extremely homogeneous in size (Figure 2) and sequence (overall sequence similarity is on average 86% for the 90% most conserved copies) suggests that their amplification took place recently during *Vitis* genome evolution. This has prompted us to analyze this MITE subfamily in more detail. In order to look for evidences of recent transposition and amplification of *mPifvine-3.1* we analyzed the presence of these elements at particular loci in 10 different cultivars of the domesticated grapevine *Vitis vinifera* ssp. *sativa*, 9 genotypes of the European wild species *Vitis vinifera* ssp. *sylvestris* and 6 more distantly related North American *Vitis* species (see Materials and Methods for details). We looked for insertion polymorphisms of *mPifvine-3.1* elements by amplifying by PCR 25 loci that contained an *mPifvine-3.1* insertion in the published genome, i.e. *Vitis vinifera* ssp. *sativa* cv. Pinot noir. Although the *mPifvine-3.1* family of MITEs is recent, pairwise comparisons of the 1298 *mPifvine-3.1* elements showed a range of sequence conservation among them allowing us to identify elements that are probably more recent than others. We have thus chosen to analyze 15 loci representing more recent insertions (shown in blue in Figure 4), as judged by their high degree of sequence similarity in pairwise comparisons, and 10 loci that probably represent older insertions, as judged by the same criterion. An example of the polymorphism analysis is shown in Figure 3, and the summary of the results obtained is presented in Figure 4. None of the *mPifvine-3.1* insertions showing very high degree of sequence similarity is present in the corresponding locus of any of the 6 American *Vitis* species, suggesting that these insertions occurred after the split of the European and American *Vitis* species. Only 11 out of 15 insertions are present in the corresponding locus for any of the European wild species of *Vitis vinifera*, and almost all of them show a high polymorphism among the 10 cultivated genotypes here tested. This confirms that these *mPifvine-3.1* insertions are recent and that *mPifvine-3.1* elements have transposed accompanying grapevine domestication and breeding. The fact

that most of these insertions are polymorphic in cultivated *Vitis vinifera* genotypes stresses the high heterozygosity of this species (Velasco *et al.* 2007). On the other hand, the 10 insertions analyzed corresponding to the less conserved *mPifvine-3.1* copies (shown in red in Figure 4) seem to be almost fixed in the population of cultivated *Vitis vinifera* genotypes and are also present in both *sylvestris* and the American *Vitis* species, confirming that these insertions are older and that *mPifvine-3.1* elements were already present and transposing in the ancestor *Vitis* species.

For a number of *mPifvine-3.1* insertion *loci*, bands of unexpected sizes were obtained (Figure 3 and Figure 4). The sequencing of bands deriving from loci 1102 and 1284 showed that these unexpected bands corresponded to deletions of the loci. In the case of locus 1102, there is a partial deletion of the *mPifvine-3.1* and its 5' flanking sequence, and in the case of locus 1284 there is a larger deletion of the whole mPifvine-3.1 and its flanking sequence. While the latter could be the result of an abortive gap repair upon *mPifvine-3.1* excision, the former does not seem to be related to the transposition of the *mPifvine-3.1* element. Regardless of the origin of the deletions, these events happened after *mPifvine-3.1* insertion, and can be used as new markers in genotyping that could provide useful information on the origin of grapevine varieties. For example it is interesting to note that the locus 1123 presents an unusual short band in four out of six American wild *Vitis* species and, although its presence in European wild *Vitis vinifera* ssp. *sylverstris* is rare (only one out of nine species has the band), it is found in domesticated grapevine varieties. Similarly, the unusual short band of locus 1284 is present in all but one European wild *Vitis vinifera* ssp. *sylverstris* while most of the *Vitis vinifera* ssp. *sativa* genotypes here analyzed do not contain this allele.

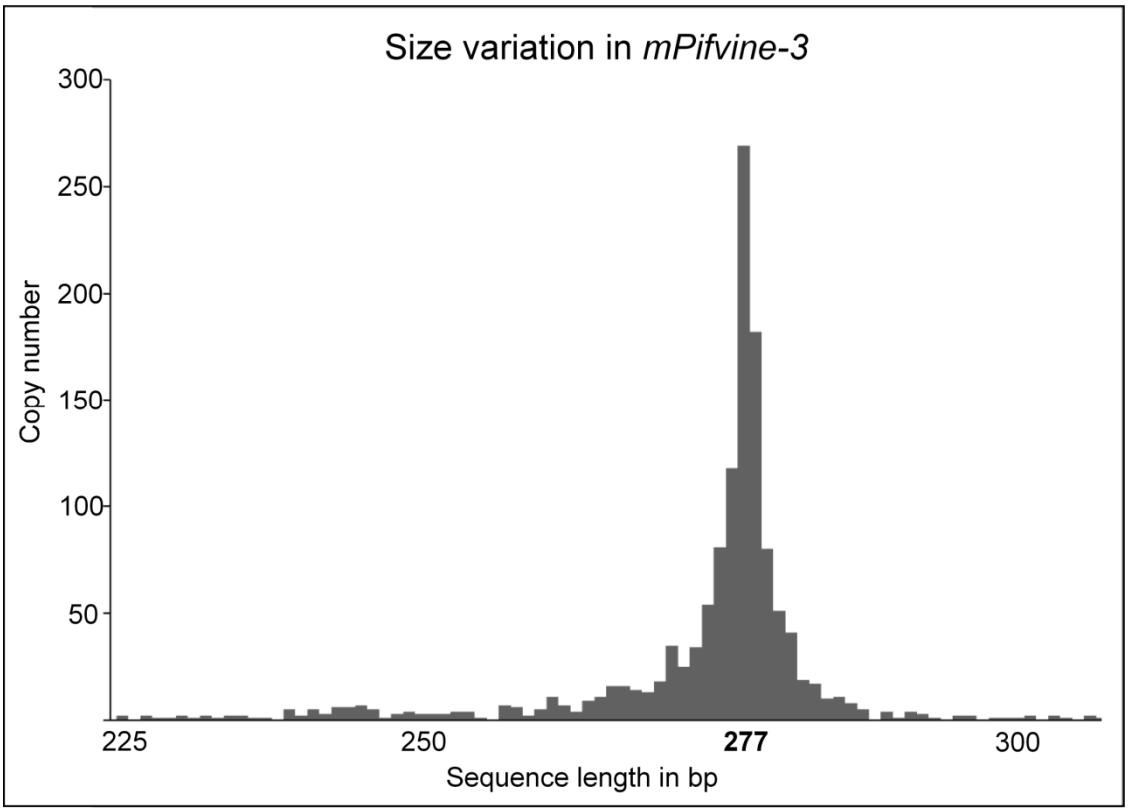

Figure 2: Size variation in *mPifvine-3.1*. Each column represents 1 nucleotide difference.

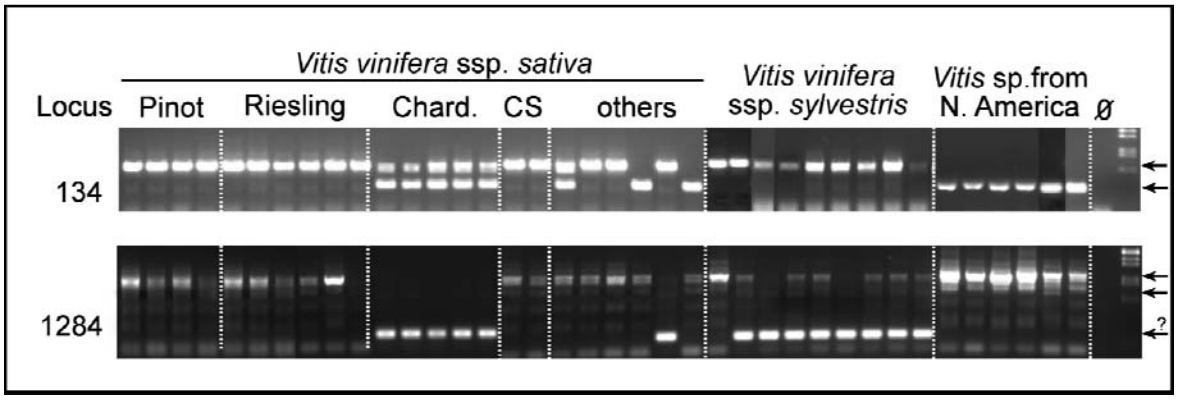

Figure 3: Examples of PCR results for two chosen loci. Pictures from different gels were joined for this figure to match the order of the samples given in Supporting Table 1. Arrows represent sizes of bands for insertion and empty sites. In the locus 1284 are only bands corresponding to the *mPifvine-3.1* insertion and unusually small bands (arrow with "?") corresponding to a larger deletion of the allele. These small bands were found in most *sylvestris* genotypes and only in 2 domesticated cultivars, Chardonnay and We 70-281-37.

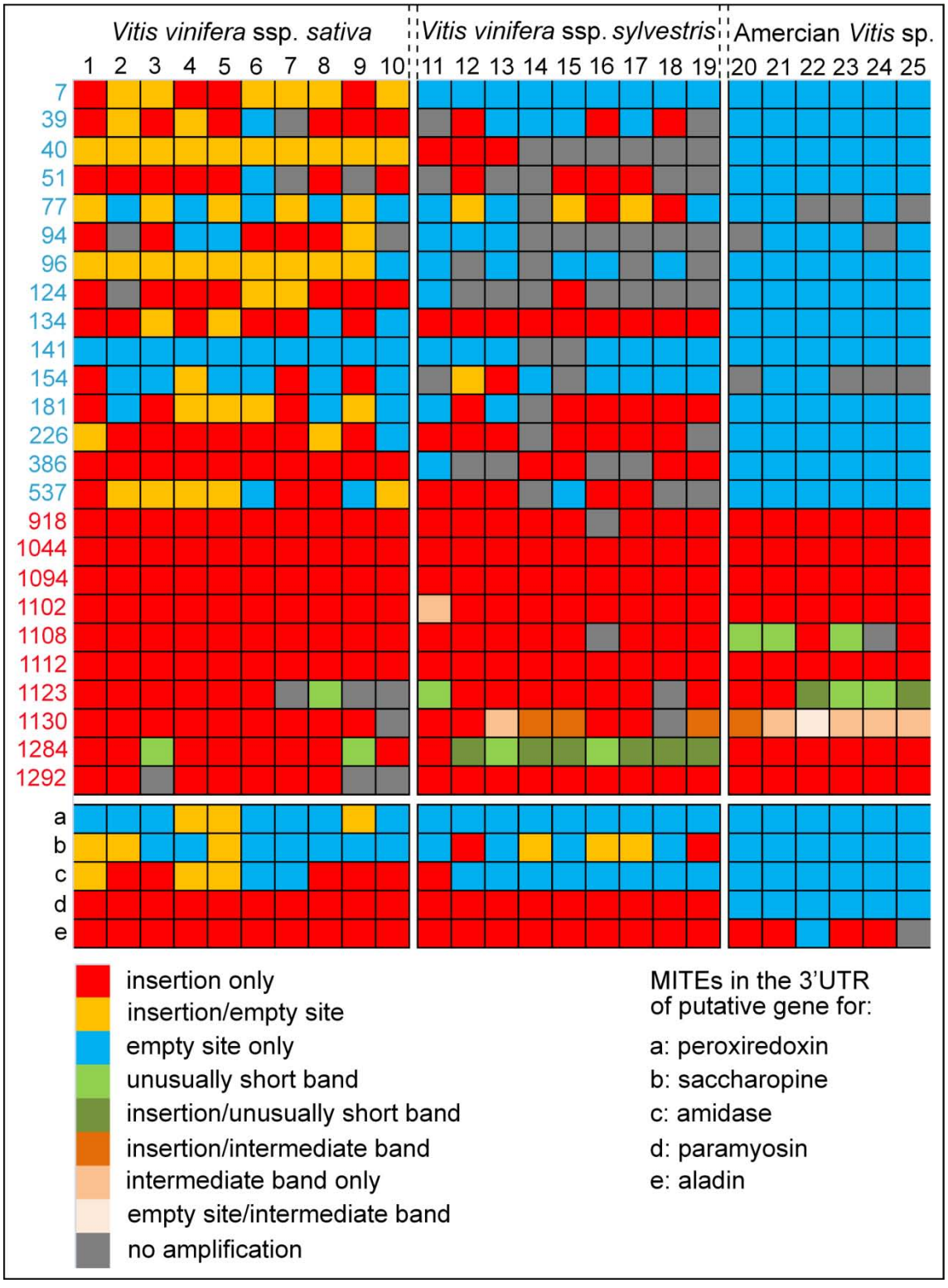

Figure 4: PCR results for *mPifvine-3.1* amplification. Samples are listed in Table 4. Numbers on the left indicate loci names and derive from the ranking of the element according to its average distance in the similarity matrix (see Materials and Methods).

mPifvine-3.1 distribution with respect to grapevine genes

MITEs are very frequently found associated to genes in plant genomes (Casacuberta and Santiago 2003, Feschotte and Pritham 2007). The potential for MITEs to generate gene variants for evolution is thus very high. We have therefore decided to analyze the possible impact of *mPifvine-3.1* amplification and transposition in the evolution of grapevine genes, by looking for the distribution of *mPifvine-3* elements with respect to *Vitis* coding sequences.

To extract the positions of all predicted genes, including intron, exon and UTR coordinates when available, we used the partially assembled grapevine chromosomes (CU462738-CU462756) which had a total of 303 Mb (~70% of the whole genome). Because we used whole genome shotgun sequences in our analysis (that allowed us to cover the whole genome), we had to *de novo* re-mine for *mPifvine-3.1* from the chromosome sequences. The simplest way to extract the positions of known MITEs from the chromosome sequences was by using the TRANSPO software (Santiago, et al. 2002) which looks for sequences of a given range of lengths and that contain a specified TIR. We used stringent parameters in order to avoid false positives, even though this lowered the efficiency of the extraction. We have extracted 498 *mPfivine-3.1*, which is close to 40% of the 1298 *mPifvine-3.1* previously found in the whole genome shotguns. We analyzed the distribution of these 498 *mPfivine-3.1* elements (that we consider to be representatives for the whole family) with respect to the predicted genes, and in cases of insertion within a predicted gene, we analyzed its possible presence in 5' UTR, exons, introns or 3' UTR. 51% of the 498 *mPifvine-3.1* elements are located in predicted genes. As gene sequences account 50% of the assembled grapevine chromosomes, our results show that there is not a bias for *mPifvine-3.1* insertion with respect to genes. Similarly, we did not find any bias of *mPifvine-3.1* distribution among chromosomes (data not shown). However, the distribution of *mPifvine-3.1* elements in the genome is not random, and certain genic regions contain an unexpectedly high number of elements inserted (Table 3).

Table 3: Analysis of the distribution of *PIF*-related MITEs in the grapevine genome.

Type of sequence	Number of MITEs/ Mb of sequence
non-genic	1.61
intron	1.49
exon	0.092
5'UTR	6.029
3'UTR	8.19

As expected, presence of mPifvine-3.1 in exons is rare, and we have found only two cases in which the insertion is partially located in a putative coding sequence. On the contrary, while the frequency of *mPifvine-3.1* insertions in introns is not far from the expected for a random distribution, the frequency of insertions into UTR is 5 to 6 fold higher than the expected. The frequent association of *mPifvine-3.1* elements with 5' and 3' UTRs, suggests that these elements should be frequently present within grapevine transcripts. We thus looked for the presence of *mPifvine-3.1* sequences in the grapevine ESTs collections available at NCBI by BLAST search. The matching ESTs were blasted back to the genome database and the genomic region was manually checked to confirm that a given MITE was the source of the EST.

We found some 50 different ESTs that match to *mPifvine-3.1* elements. In most cases the insertion is found within the predicted 3'UTR of the transcript. Interestingly, in some cases different ESTs corresponding to the same gene are polymorphic with respect to the presence of the MITE (Figure 5). This can be the result of an insertion polymorphism among different cultivars (as the EST collections contain sequences obtained from different cultivars) or can be due to the existence of two alleles of the same gene in a particular cultivar. For example, two different alleles for a putative isoflavone reductase (CAN80172.1) were found in Pinot noir, one of which has an *mPifvine-3* insertion at the 3' end of the gene (Figure 5, f), suggesting that Pinot noir has two alleles of the gene, only one of them containing the *mPifvine-3.1* insertion, and that both alleles are transcribed. Indeed, transcripts for both alleles were found in ESTs deriving from Shiraz and Chardonnay, but Cabernet Sauvignon, Muscat Hamburg and Thompson Seedless only had ESTs with the *mPifvine-3.1* insertion. Similarly, we found different ESTs corresponding to a gene coding for

a putative saccharopine dehydrogenase (CAO15039.1) with or without an *mPifvine-3.1* element inserted in the 3'UTR. In this case, only one EST was found with the *mPifvine-3.1* insertion and 9 ESTs with the empty site, suggesting that the transcript containing the insertion is expressed at a lower level.

We have selected five *mPifvine-3.1* insertions present in grapevine EST collections to further analyze their insertion polymorphisms among 25 different *Vitis* genotypes. All these insertions generated new transcription termination sites for the genes where they inserted in, as deduced from the analysis of ESTs collections. One of these insertions (designated as "e" in Figures 4 and 5) is probably an old insertion that occurred prior to the split between European and American *Vitis* species and is almost fixed in the species analyzed (only one American species does not contain it). Still, the grapevine EST collections contain two different transcripts corresponding to the gene where this *mPifvine-3.1* is inserted, one of them stopping just before the MITE sequence. The second insertion analyzed, designated with a "d" in Figures 4 and 5, is absent from all the American wild species analyzed, while it seems to be fixed in the European wild and domesticated species. The rest of the insertions analyzed are highly polymorphic among domesticated *Vitis* species. Two of them ("b" and "c" in Figures 4 and 5) are also present in some European wild species suggesting that the insertion occurred before grapevine domestication, while the third ("a" in Figures 4 and 5), seems to be specific of the domesticated genotypes suggesting that it occurred accompanying grapevine domestication and breeding. In all these cases, the grapevine EST collections contain two different transcripts corresponding to the two alleles found in grapevine genotypes, showing that the insertion of the *mPifvine-3.1* element has generated transcript variability.

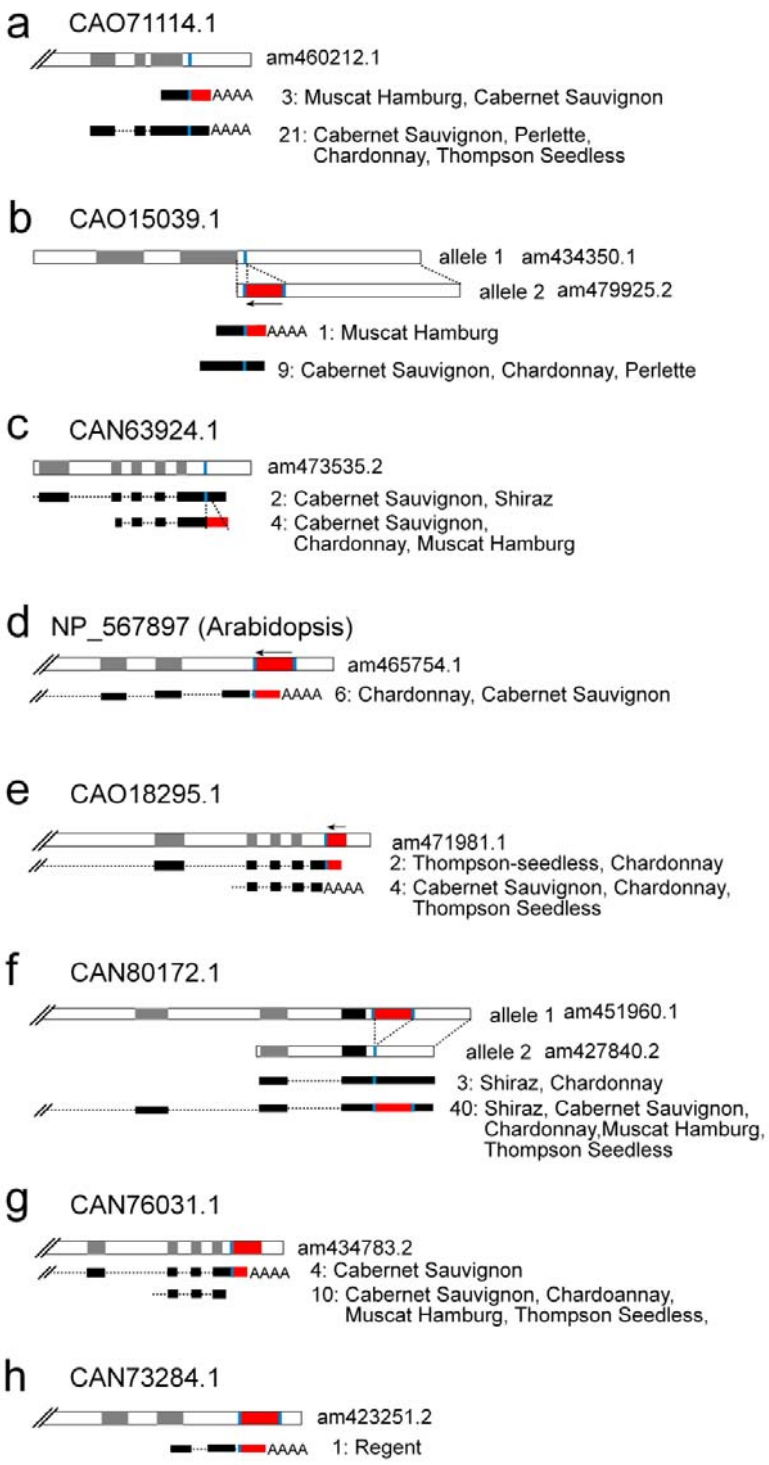

Figure 5: Presence of the *m-Pifvine-3.1* MITEs (in red) in transcripts of different genes. TSDs are in blue. White lines represent genes and black lines are ESTs. All genes are 5'-3' orientated. Grey boxes are exons (for some genes not all exons are shown). Arrows indicate cases where MITEs are in reverse orientation. Accession numbers of genomic sequences are given as well as accessions of predicted proteins corresponding to the genes. "AAAA" means polyadenylation site.

Discussion

While the first MITEs described were related to the *PIF*/*Pong* and *Mariner* families, it has been shown later on that most class II families of TEs can generate MITEs. Here we show that grapevine also contains MITEs related to most of class II TEs families present in this species. It has been proposed that MITEs are generated by a two-step process in which a subset of defective class II elements with special characteristics (e.g. small size) would be amplified to high copy numbers by a replicative-related, and still to be described, mechanism (Casacuberta and Santiago 2003, Feschotte, et al. 2002). Such a mechanism implies that MITEs are amplified from typical class II elements and thus, both types of elements should co-exist in a particular genome. This is what has been found for the *impala*/*mimp1* element of *Fusarium oxisporium* (Dufresne *et al.* 2007), but in other cases, such as that of the Arabidopsis and *Medicago truncatula Emigrant*/*Lemi1* elements, MITEs and typical defective elements are restricted to different genomes (Guermonprez *et al.* 2008). The work presented here shows that defective elements and MITEs do co-exist in grapevine and, more significantly, that grapevine also contains elements that could represent an intermediate type of defective elements. Indeed, seven out of the eight families here described contain elements that, while being highly homogeneous in size and sequence, are relatively long and are present at moderate copy number. This new type of defective class II elements could be the result of an incomplete amplification due to sub-optimal characteristics of the family founder element. In this respect, it is interesting to note the inverse relationship between the size of the elements and their copy number, suggesting that the size could be an important constraint for the high amplification of a defective element. The results presented here thus support a model for MITEs amplification from particular defective class II elements and point to a small size as one of the important characteristic for a defective element to become the founder of a new MITE family.

While the elements described here were found because they show extensive sequence similarity with the related grapevine TEs families, some of the longest elements also contain internal sequences not related to them but to grapevine genomic sequences. Transposons can capture and mobilize genome sequences, and we have recently shown that this phenomenon also occurred in grapevine (Benjak, et al. 2008). But, to our knowledge, the

capacity of MITEs to transduplicate genomic sequences has not been reported to date. Transposons are usually present at low or moderate copy numbers, and with the exception of the Arabidopsis KI-MULE that is present in some 97 copies (Hoen *et al.* 2006), the transduplicated sequences are unique or at present at very low copy number. It has been proposed that transduplicated gene fragment may regulate paralogous gene expression through siRNA-related mechanisms, or they may provide sequence reservoirs for gene conversion (Hoen, et al. 2006). The capacity of MITEs to transduplicate genome sequences greatly increases the possibility of amplification and mobilization of transduplicated gene fragments and may have important implications for the evolution and regulation of the related genes.

In addition to genic sequences, grapevine MITEs have also transduplicated a fragment of a previously uncharacterized transposon that we have named *Mila*. The amplification within a MITE of a transduplicated transposon fragment will increase the possibilities for a siRNA control of the transposon and may represent a new mechanism to control transposon activity.

One of the MITE families described here, *mPifvine-3.1*, has attained more than 1000 copies in grapevine. The low sequence variability of *mPifvine-3.1* elements suggests that they have been amplified recently during grapevine evolution. Our results show that, although *mPifvine-3.1* were already present in the ancestor of the wild *Vitis* species found in both Europe and America, they have transposed and amplified after their split accompanying grapevine domestication and breeding. More than half of the *mPifvine-3.1* are closely associated to grapevine genes and they are frequently located within the 3' UTR sequence. These insertions are frequently present in grapevine transcripts where they might influence its fate in many ways, including its stability and processing or its degradation through post-transcriptional gene silencing mechanisms. It is interesting to note that, while older *mPifvine-3.1* insertions are fixed in the population, the recent ones are highly polymorphic among cultivars. This polymorphism, which can be detected also at the transcriptional level, may be linked to phenotypic variability.

Although sexual crossing has been a major driver of grapevine evolution, its vegetative propagation enhanced the impact of somatic mutations and has been important for

grapevine diversity. Transposable elements are known to be major contributors to genome variability and, in particular, to somatic mutations. Among them, MITEs seem particularly well suited to influence gene evolution. Their smaller size may allow MITEs to introduce more subtle changes in gene expression or regulation, and their high copy number makes their potential impact higher.

In addition to the fundamental interest for genome evolution studies, MITE insertion can be also used as molecular markers. The work presented here shows that *mPifvine-3.1* "insertion/empty site" bands represent co-dominant alleles at a single locus that can be used for fingerprinting. The high copy number of *mPifvine-3.1* elements as well as their frequent association to genes make them a very useful potential source for new marker to assist selection programs as well as for varietal and clone identification. It has to be noted that, although we have analyzed in detail only one subfamily of *mPifvine-3* elements, during our work we have also detected four new subfamilies of the *mPifvine-3* MITEs (that share only the TIR sequence with the *mPifvine-3.1*) and that in total account for additional 2000 elements (not shown).

In summary, the work presented here shows that MITEs have contributed to gene evolution in grapevine by capturing and amplifying gene sequences as well as by inserting in a high number of grapevine genes.

Materials and methods

Transposon mining

We performed our analyses using the whole genome shotgun sequences of the two sequenced grapevine genomes made available at NCBI (Jaillon*, et al.* 2007, Velasco*, et al.* 2007). We retrieved the putative MITEs by blasting (Altschul *et al.* 1990) previously described TEs (Benjak*, et al.* 2008). To check for transcription of MITEs, representatives for each MITE family were blasted against the grapevine EST collection at NCBI. The matching ESTs were blasted back to the nucleotide database to determine the source sequence for each transcript. As both Velasco et al. (2007) and Jaillon et al. (2007) performed

computational gene predictions, the NCBI contains a significant number of predicted (but not annotated) *Vitis* proteins which were useful to characterize the transduplicated sequences. For each MITE group, multiple alignments were created from which similarity matrices were calculated using BioEdit software (Hall 1999). Average similarities were calculated from similarity matrices in Microsoft Excel software.

Plant material

A list of samples and their source is given in Table 4. DNA from all samples derived from Germany was extracted using E.Z.N.A. SP Plant DNA Mini Kit (Omega Bio-tek). DNA of other samples was obtained from different laboratories.

Table 4: List of samples used for the PCR analysis and their source

Vitis species	Nº	Cultivar	Clone/accession	Origin
V. v. sativa	1	Pinot	We 242	LVWO Weinsberg, Germany
			We 111	
			20 Gm	
			1-84 Gm	
	2	Riesling	198-44 Gm	Institute of Grapevine Breeding Geisenheim , Germany
			64-183 Gm	
			110-14 Gm	
			24-195 Gm	
			239-17 Gm	
		R. Riesling	23 Gm	
	3	Chardonnay	50 Gm	
			1 Gm	
			33 Gm	
			3 Gm	
			52 Gm	
	4	Cabernet Sauvignon	Levadoux	LVWO Weinsberg, Germany
			Gm 1	
	5	Cabernet Mitos	-	Rebveredler Antes, Heppenheim, Germany
	6	Cabernet Cortis	-	
	7	Cabernet Carbon	-	
	8	Cabernet Dorsa	We 750	
	9	Lemberger x Cab. Sau.	We 70-281-37	LVWO Weinsberg, Germany
	10	Acolon		
V. v. sylvestris	11	-	CA4-4	INIA, Madrid, Spain
	12	-	S 47-7	
	13	Grésigne-1	Sample V6037a	INRA, Montpellier, France
	14	-	N1VS	BOKU, Vienna, Austria (samples derive from the area of Germany and Switzerland)
	15	-	N6BVS	
	16	-	N26AVS	
	17	-	N29AVS	
	18	-	N31VS	
	19	-	N32AVS	
V. rufotomentosa	20	-	DVIT 1416	Department of Viticulture and Enology, UC Davis, USA
V. champinii	21	-	DVIT 9036	
V. cinerea	22	Barrett 9	DVIT 1363	
V. riparia	23	Riparia Gloire	DVIT 1437	
V. ruperstris	24	St. George	UCD2A	
V. riparia	25	Michaux	-	LVWO Weinsberg

PCR analysis

Primers were designed using Primer3 (Rozen and Skaletsky 2000) and FastPCR programs (Kalendar 2006, www.biocenter.helsinki.fi/bi/programs/fastpcr.htm). Each primer was blasted against the whole *Vitis* genomic database to check for specificity. The list of primers is given in Supporting Table 1. PCRs were done in 20 µl reaction volumes using approximately 30 ng of DNA template, 0.5 µl of each primer (10 pmol/µl), and TaKaRa Ex Taq in the following conditions: 94 ºC·2min^{-1} + 40 x (94 ºC·25s^{-1}, 58-62 ºC (depending on primer)·45s^{-1}, 72 ºC·1min^{-1}) + 72 ºC·5min^{-1}. PCR products were run in 1.2% agarose gels with EtBr in a 1x TAE buffer and visualized under UV light.

Gene MITE positions extraction and analysis

We have extracted the positions of all predicted genes, including intron, exon and UTR coordinates when available from the grapevine chromosomes CU462738- CU462756 using a Perl script. To mine for MITEs, we used the TRANSPO 1.0 software (http://alggen.lsi.upc.es/) and stringent parameters in order to avoid false positives, even though lowering the efficiency of the extraction. The number of mismatches allowed for the TIR sequences in the TRANSPO program was 1 for the *mPifvine-3.1*. The positions of MITEs were extracted from TRANSPO results and compared to the positions of genes using a Perl script.

Acknowledgements

For providing the plant material we thank Ernst Rühl (Institute of Grapevine Breeding Geisenheim, Germany), Bernd Hill (LVWO Weinsberg, Germany), and Reinhard Antes (Nursery Antes, Heppenheim, Germany). For providing DNA material we thank Prof. Andrew Walker and Summeira Riaz (Department of Viticulture and Enology, UC Davis, USA), Patrice This (INRA, Montpellier, France) and Jose-Miguel Martínez-Zapater and Rosa Arroyo-García (INIA, Madrid, Spain). This work was partially funded by a grant of the Ministerio de Educación y Ciencia (Grant BFU2006-04005) and the Xarxa de Referència en Biotecnologia

from the Generalitat de Catalunya to J.M.C. A.B. was partially funded by an EMBO Short Term Fellowship (ASTF 115-07).

References

Benjak, A., Forneck, A. and Casacuberta, J.M. (2008) Genome-Wide Analysis of the "Cut-and-Paste" Transposons of Grapevine. *PLoS ONE*, **3**, e3107.

Casacuberta, J.M. and Santiago, N. (2003) Plant LTR-retrotransposons and MITEs: control of transposition and impact on the evolution of plant genes and genomes. *Gene*, **311**, 1-11.

Deragon, J., Casacuberta, J.M. and Panaud, O. (2008) Plant Transposable Elements. In *Genome Dyn* (J-N, V. ed. Basel: Karger, pp. 69-82.

Dufresne, M., Hua-Van, A., Abd el Wahab, H., M'Barek, S.B., Vasnier, C., Teysset, L., Kema, G.H.J. and Daboussi, M.-J. (2007) Transposition of a Fungal Miniature Inverted-Repeat Transposable Element Through the Action of a Tc1-Like Transposase. *Genetics*, **175**, 441-452.

Feschotte, C., Jiang, N. and Wessler, S.R. (2002) Plant transposable elements: where genetics meets genomics. *Nat Rev Genet*, **3**, 329-341.

Feschotte, C. and Mouches, C. (2000) Evidence that a Family of Miniature Inverted-Repeat Transposable Elements (MITEs) from the Arabidopsis thaliana Genome Has Arisen from a pogo-like DNA Transposon. *Mol Biol Evol*, **17**, 730-737.

Feschotte, C. and Pritham, E.J. (2007) DNA Transposons and the Evolution of Eukaryotic Genomes. *Annual Review of Genetics*, **41**, 331-368.

Feschotte, C., Swamy, L. and Wessler, S.R. (2003) Genome-Wide Analysis of mariner-Like Transposable Elements in Rice Reveals Complex Relationships With Stowaway Miniature Inverted Repeat Transposable Elements (MITEs). *Genetics*, **163**, 747-758.

Guermonprez, H., Loot, C. and Casacuberta, J.M. (2008) Different Strategies to Persist: The pogo-Like Lemi1 Transposon Produces Miniature Inverted-Repeat Transposable Elements or Typical Defective Elements in Different Plant Genomes. *Genetics*, **180**, 83-92.

Hoen, D.R., Park, K.C., Elrouby, N., Yu, Z., Mohabir, N., Cowan, R.K. and Bureau, T.E. (2006) Transposon-Mediated Expansion and Diversification of a Family of ULP-like Genes. *Mol Biol Evol*, **23**, 1254-1268.

Jiang, N., Feschotte, C., Zhang, X. and Wessler, S.R. (2004) Using rice to understand the origin and amplification of miniature inverted repeat transposable elements (MITEs). *Current Opinion in Plant Biology*, **7**, 115-119.

Juretic, N., Hoen, D.R., Huynh, M.L., Harrison, P.M. and Bureau, T.E. (2005) The evolutionary fate of MULE-mediated duplications of host gene fragments in rice. *Genome Res.*, **15**, 1292-1297.

Kuang, H., Padmanabhan, C., Li, F., Kamei, A., Bhaskar, P.B., Ouyang, S., Jiang, J., Buell, C.R. and Baker, B. (2008) Identification of miniature inverted-repeat transposable elements (MITEs) and biogenesis of their siRNAs in the Solanaceae: New functional implications for MITEs. *Genome Research*, -.

Lepetit, D., Pasquet, S., Olive, M., Theze, N. and Thiebaud, P. (2000) Glider and Vision: two new families of miniature inverted-repeat transposable elements in Xenopus laevis genome. *Genetica*, **108**, 163-169.

Moreno-Vázquez, S., Ning, J. and Meyers, B. (2005) hATpin, a Family of MITE-like hAT Mobile Elements Conserved in Diverse Plant Species that Forms Highly Stable Secondary Structures. *Plant Molecular Biology*, **58**, 869-886.

Ohmori, Y., Abiko, M., Horibata, A. and Hirano, H.-Y. (2008) A Transposon, Ping, is Integrated into Intron 4 of the DROOPING LEAF Gene of Rice, Weakly Reducing its Expression and Causing a Mild Drooping Leaf Phenotype. *Plant Cell Physiol.*, **49**, 1176-1184.

Piriyapongsa, J. and Jordan, I.K. (2007) A Family of Human MicroRNA Genes from Miniature Inverted-Repeat Transposable Elements. *PLoS ONE*, **2**, e203.

Piriyapongsa, J. and Jordan, I.K. (2008) Dual coding of siRNAs and miRNAs by plant transposable elements. *RNA*, **14**, 814-821.

Saito, M., Yonemaru, J., Ishikawa, G. and Nakamura, T. (2005) A candidate autonomous version of the wheat MITE Hikkoshi is present in the rice genome. *Molecular Genetics and Genomics*, **273**, 404-414.

Santiago, N., Herraiz, C., Goni, J.R., Messeguer, X. and Casacuberta, J.M. (2002) Genome-wide Analysis of the Emigrant Family of MITEs of Arabidopsis thaliana. *Mol Biol Evol*, **19**, 2285-2293.

Velasco, R., Zharkikh, A., Troggio, M., Cartwright, D., Cestaro, A., Pruss, D., Pindo, M., Fitzgerald, L., Vezzulli, S., Reid, J., Malacarne, G., Iliev, D., Coppola, G., Wardell, B., Micheletti, D., Macalma, T., Facci, M., Mitchell, J., Perazzolli, M., Eldredge, G., Gatto, P., Oyzerski, R., Moretto, M., Gutin, N., Stefanini, M., Chen, Y., Segala, C., Davenport, C., Demattè, L., Mraz, A., Battilana, J., Stormo, K., Costa, F., Tao, Q., Si-Ammour, A., Harkins, T., Lackey, A., Perbost, C., Taillon, B., Stella, A., Solovyev, V., Fawcett, J., Sterck, L., Vandepoele, K., Grando, S., Toppo, S., Moser, C., Lanchbury, J., Bogden, R., Skolnick, M., Sgaramella, V., Bhatnagar, S., Fontana, P., Gutin, A., Van de Peer, Y., Salamini, F. and Viola, R. (2007) A High Quality Draft Consensus Sequence of the Genome of a Heterozygous Grapevine Variety. *PLoS ONE*, **2**, e1326.

Yang, G. and Hall, T.C. (2003) MDM-1 and MDM-2: Two Mutator-Derived MITE Families in Rice. *Journal of Molecular Evolution*, **56**, 255-264.

Zhang, X., Jiang, N., Feschotte, C. and Wessler, S.R. (2004) PIF- and Pong-Like Transposable Elements: Distribution, Evolution and Relationship With Tourist-Like Miniature Inverted-Repeat Transposable Elements. *Genetics*, **166**, 971-986.

DISCUSSION

A WORD OR TWO ABOUT THE ARTICLES

The article 2 was the most methodological. It highlights the important issue of the DNA quality to be used in digestion-based and PCR-based fingerprinting. DNA from *Vitis sp.* is usually more difficult to extract compared to other plants like *Arabidopsis*, tobacco or wheat (personal experience) and is often less pure (Lodhi *et al.* 1994). This might have consequences even in robust analytical methods. For example, specific target PCRs are usually less sensitive to the DNA quality, but could still be affected. This was the case in the work presented in articles 5 and 6 where triple DNA extractions were done from the same plants, always from young leaves and using the same DNA extraction kit. The DNA yields were similar. In general, too much DNA template (1 µl of the original DNA elution for 20-25 µl of PCR reaction) would inhibit the PCR reaction in many cases. When template reduced (0.5 µl), PCR performed well, but would still not work for some DNA repetitions. Considering the PCR problems that can occur in specific target amplification (where the PCR conditions are stringent, and primer sites are relatively less abundant) one cannot avoid to think about AFLP-based fingerprinting techniques which are more sensitive, genome wide and which produce a high number of bands. The article 2 is not criticizing the AFLP but it suggests an extra care when applying this method for the analysis of clonal variability where only a small number of variable bands are detected among a big number of total amplified bands.

Article 3 was an attempt to implement universal retrotransposon primers into the S-SAP fingerprinting method. This work was done prior to the availability of the draft genome sequence. Even though the universal primers used for this work are now less helpful in the light of the genomic sequence available along with a number of retrotransposon sequences already annotated and uploaded to the Repbase database, the value of the work consist on underlining the idea of using retrotransposons as variable genomic elements for detection of clonal variability. For this idea to work more efficiently it is necessary to not only annotate, but also fully characterize all retrotransposons in the grapevine genome. The high copy number elements which retained the ability to transpose are better candidates to be

used in the studies of clonal variability. In this regard, article 4 went a step further. In this work, the first step was to annotate a number of retrotransposons. Then, each family of the annotated elements was independently used for fingerprinting by designing specific primers. The strategy used in article 4 was logical but methodologically delicate for the first step, the annotation of retrotransposons. Genome-wide annotation of TEs is generally complex (reviewed in Bergman and Quesneville 2007). Moreover, after additional checkup some LTR groups turned out to belong to one family of elements which consisted of higher number of slightly divergent sequences. The annotation of LTR-retrotransposons done in article 4 was clearly not complete but provided preliminary and valuable information on the LTR content of the grapevine genome. While mining putative LTR sequences does not appear to be a problem, the tricky part is to characterize the elements found. Without this characterization the practical usage of LTR primers and interpretation of the result is very limited. The perfect optimization of this method would include a genome wide characterization of retrotransposons and their comparative analysis. Once all the sequences are collected, annotated and aligned (which would be a substantial amount of work), it would be possible to design primers specific only to certain elements of interest, like the highly abundant ones or potentially active, for example. Only then, the IRAP, REMAP or possible derivates of these methods, could be used for genetic analysis of grapevine, with more focus on retrotransposon activity and mutagenesis, rather than general fingerprinting, for which different methods already exist.

Articles 5 and 6 focus entirely on class II TEs in grapevine. The goal of article 5 was to characterize in detail all “copy-paste”-type of class II TEs. Article 6 represents a continuation of the previous one by analyzing more in detail some MITE families. It is not the report on abundance and copy number of class II TEs but rather the characterization of some of their extraordinary features that gives a great biological significance to these works. These are gene transduplication, TE domestication, insertion polymorphism and transcription. While the article 5 points more toward the evolutionary impact of TEs on the grapevine genome, article 6 stresses the idea that MITEs might directly shape the grapevines phenotype and contribute to genomic variability. This theory is supported by the polymorphism found among grapevine cultivars for insertions of MITEs into or close to genes, presence of MITEs

in the UTRs of some gene transcripts and recent MITE insertions that are present only in the genome of *Vitis vinifera* ssp. *sativa*.

A BROADER PERSPECTIVE

The sequencing of the grapevine genome brings the grapevine genetics to a higher level of research. Genomic physical maps are being constructed (Moroldo *et al.* 2008, Vezzulli *et al.* 2008), putative genes predicted (Jaillon *et al.* 2007, Velasco *et al.* 2007), mitochondrial and chloroplast DNA analyzed (Goremykin *et al.* 2008), genome-wide SNP analyses being done (Pindo *et al.* 2008) and so on. It is now easier and more efficient to identify and comparatively analyze genes (Abbal *et al.* 2008, Chong *et al.* 2008, Costantini *et al.* 2008, Nonis *et al.* 2008, Xiao *et al.* 2008). But not only is the genomic study enhanced. More work is being done recently in transcriptomics (Doddapaneni *et al.* 2008, Iandolino *et al.* 2008) and grapevine proteomics (Marsoni *et al.* 2008). TEs play a big part in all this because of their impact on different levels of the cell machinery. TEs change genomic DNA sequence, modify the chromosome structure, influence gene regulation and their expression and even contribute to the proteome of an organism (see details in the Introduction). With today's knowledge it is clear that there is no big difference in the significance of TEs and, for example genes or non-coding RNA. The term "host genome" is still widely used, more as a historical leftover (even I use it sometimes), but it is actually wrong. The term comes from the earlier idea that TEs evolved separately from an organism and, as a disease, invaded every genome on this planet (or more likely only the ancestral genome) and now continue to proliferate in these genomes. But from the DNA point of view and its evolution, this does not really make sense. It could well also be genome in the "host transposons". Some scientist might suggest that TEs are an inevitable consequence of the DNA replication and its way to survive, like an evolutionary side effect. But are not also genes as well "side effects" of the chemical property of the DNA replication? Genomes obviously have advantages from having "developed" TEs, if not TEs would be already erased from genomes by the selective pressure during the DNA evolution. On the contrary, TEs evolved very early as they are found in most primitive organisms, as are genes. The term "gene" is constantly changing, from previously thought as DNA stretches that encode for proteins, to a much more

complex perspective of the genomic machinery as suggested by Gerstein *et al.* 2007, "a gene is a union of genomic sequences encoding a coherent set of potentially overlapping functional products".

CAN YOU SEE THE HORIZON?

In the perspective described above, TEs will be analyzed in the grapevine genome in the context of the complex genomic and functional system of a living cell. In my opinion, we are still miles away from deciphering the way a living cell works. More and more genomic data of different kind are collected and tried to be linked together in a logic way in order to see patterns, or any kind of rule linked to a biological function. Microarrays are good example where a big amount of data is analyzed in order to get some clues on what is going on in the living cell. Complexity is the major limitation in understanding the living processes because of a virtually infinite number of possible interactions among DNA, RNA and proteins.

A big part of the research in the grapevine genomics in the past decades was done in the field of DNA fingerprinting, which more or less consisted on developing a practical way to detect differences in the genomic DNA sequence among individuals. Methods that target specific loci, like SSRs, will be always useful because they are simple to analyze, but are restricted for a limited types of analyses. On the other hand, genome-wide fingerprinting assays are meant to detect as much differences as possible among genomes. The currently most used AFLP-based methods will soon be overcome by whole genome sequencing. Currently, the most advanced sequencing method, the "454 sequencing" (named after the 454 Life Sciences company owned by Roche) is capable of sequencing roughly 400-600 megabases of DNA per 10-hour run on the Genome Sequencer FLX with GS FLX Titanium series reagents. This means that the whole grapevine genome could be sequenced in 1 day! Of course, much more time is needed thereafter to correctly assemble and analyze the sequences. Still, considering that this technology is only couple of years old, it is very likely that in the next few years sequencing and sequence assembly will be even faster. This will enable scientist to routinely sequence whole genomes from different samples in a very short time. It does not require too much imagination to even think of monitoring the whole-

genome sequence in real time. It will be a huge amount of data to process, and scientists will spend more time behind computers than pipetting samples in the lab. Even then, efficient sequencing would only enable us to see differences in DNA sequences while the idea of fingerprinting goes further than this. Fingerprinting must help discovering functional units in the genome (remember that genes are complex unions of coding sequences and all regulatory elements that may affect them, including TEs), or link genes to a certain trait. In other words, fingerprinting must help discovering genetic markers. This will contribute to the understanding of the genomic processes which will in turn be used in plant breeding or biotechnological engineering.

ACKNOWLEDGMENTS

I started to show interest in grapevine genomics during my bachelor studies at the Faculty of Agriculture, University of Zagreb. I am grateful to Prof. Dr. Ivan Pejić who introduced me to science and integrated me in several of his research projects including international collaborations. By the end of my diploma study, he introduced me to Prof. Dr. Astrid Forneck, who was looking for PhD students. That is how my PhD started. From the first day, Astrid gave me absolute freedom and support to pursue my own scientific ideas. I am very grateful for that, as this is the biggest motivation that a PhD student can have. Away from the lab, Astrid continues to be a very good friend, which is very important to me.

I wish to thank my colleagues and friends at BOKU. I have spent two very nice years in Vienna (Manuel, du bekommst ein extra Bierkasten für deine Hilfe, wie versprochen). Special thanks to our secretaries who always promptly helped with the administrative issues.

I am thankful to Dr. Ruslan Kalendar and Prof. Dr. Alan Schulman from the Institute of Biotechnology at the University of Helsinki, who showed great interest in collaborative projects and who kindly welcomed me in their lab. It was a lot of fun.

The last year and a half I have spent in the lab of Dr. Josep (Pep) Casacuberta in Barcelona. Pep is extraordinary supportive and friendly and it is such a great pleasure to work and to spend time with him. As a consequence, the atmosphere in his lab is the best one can imagine, thanks also to the nice people working there: ¡Gracias a todas las chicas del laboratiorio!

Apart from work, my private life changed quite a lot since my arrival in Barcelona. Stéphanie and I got married, we moved to a house outside the city and few months later my daughter Mila was born. Sounds like a fairytale, doesn't it? They will always be my absolute support and motivation.

In conclusion, I want to thank everybody that helped me and was close to me throughout this time, from my family to my friends and colleagues.

Andrej

REFERENCES

These references correspond to the Introduction and Discussion of the thesis.

Abbal, P., Pradal, M., Muniz, L., Sauvage, F.X., Chatelet, P., Ueda, T. and Tesniere, C. (2008) Molecular characterization and expression analysis of the Rab GTPase family in Vitis vinifera reveal the specific expression of a VvRabA protein. *J Exp Bot*, 59, 2403-2416.

Aradhya, M.K., Dangl, G.S., Prins, B.H., Boursiquot, J.M., Walker, M.A., Meredith, C.P. and Simon, C.J. (2003) Genetic structure and differentiation in cultivated grape, Vitis vinifera L. *Genet Res*, 81, 179-192.

Arnold, C., Schnitzler, A., Douard, A., Peter, R. and Gillet, F. (2005) Is there a future for wild grapevine (*Vitis vinifera* subsp. *silvestris*) in the Rhine Valley? *Biodiversity and Conservation*, 14, 1507-1523.

Arroyo-Garcia, R., Ruiz-Garcia, L., Bolling, L., Ocete, R., Lopez, M.A., Arnold, C., Ergul, A., Soylemezoglu, G., Uzun, H.I., Cabello, F., Ibanez, J., Aradhya, M.K., Atanassov, A., Atanassov, I., Balint, S., Cenis, J.L., Costantini, L., Goris-Lavets, S., Grando, M.S., Klein, B.Y., McGovern, P.E., Merdinoglu, D., Pejic, I., Pelsy, F., Primikirios, N., Risovannaya, V., Roubelakis-Angelakis, K.A., Snoussi, H., Sotiri, P., Tamhankar, S., This, P., Troshin, L., Malpica, J.M., Lefort, F. and Martinez-Zapater, J.M. (2006) Multiple origins of cultivated grapevine (Vitis vinifera L. ssp. sativa) based on chloroplast DNA polymorphisms. *Mol Ecol*, 15, 3707-3714.

Bartel, D.P. (2004) MicroRNAs: genomics, biogenesis, mechanism, and function. *Cell*, 116, 281-297.

Benjak, A., Forneck, A. and Casacuberta, J.M. (2008) Genome-wide analysis of the "cut-and-paste" transposons of grapevine. *PLoS ONE*, 3, e3107.

Bennetzen, J.L. (2000) Transposable element contributions to plant gene and genome evolution. *Plant Molecular Biology*, 42, 251-269.

Bennetzen, J.L., Ma, J.X. and Devos, K.M. (2005) Mechanisms of recent genome size variation in flowering plants. *Annals of Botany*, 95(1), 127-132.

Bergman, C.M. and Quesneville, H. (2007) Discovering and detecting transposable elements in genome sequences. *Brief Bioinform*, 8, 382-392.

Bowers, J., Boursiquot, J.-M., This, P., Chu, K., Johansson, H. and Meredith, C. (1999) Historical Genetics: The Parentage of Chardonnay, Gamay, and Other Wine Grapes of Northeastern France. *Science*, 285, 1562-1565.

Boyko, A. and Kovalchuk, I. (2008) Epigenetic control of plant stress response. *Environ Mol Mutagen*, 49, 61-72.

Buchon, N. and Vaury, C. (2006) RNAi: a defensive RNA-silencing against viruses and transposable elements. *Heredity*, 96, 195-202.

Bundock, P. and Hooykaas, P. (2005) An Arabidopsis hAT-like transposase is essential for plant development. *Nature-London*, 436(7048), 282-284.

Bureau, T.E., White, S.E. and Wessler, S.R. (1994) Transduction of a cellular gene by a plant retroelement. *Cell*, 77, 479-480.

Chong, J., Le Henanff, G., Bertsch, C. and Walter, B. (2008) Identification, expression analysis and characterization of defense and signaling genes in Vitis vinifera. *Plant Physiol Biochem*, 46, 469-481.

Costa, F.F. (2008) Non-coding RNAs, epigenetics and complexity. *Gene*, 410, 9-17.

Costantini, L., Battilana, J., Lamaj, F., Fanizza, G. and Grando, M.S. (2008) Berry and phenology-related traits in grapevine (Vitis vinifera L.): from quantitative trait loci to underlying genes. *BMC Plant Biol*, 8, 38.

Cowan, R.K., Hoen, D.R., Schoen, D.J. and Bureau, T.E. (2005) MUSTANG is a Novel Family of Domesticated Transposase Genes Found in Diverse Angiosperms. *Molecular Biology and Evolution*, 22, 2084-2089.

Dietsch, M.-F. (1996) Gathered fruits and cultivated plants at Bercy (Paris), a Neolithic village in a fluvial context. *Vegetation History and Archaeobotany*, 5, 89-97.

Doddapaneni, H., Lin, H., Walker, M.A., Yao, J. and Civerolo, E.L. (2008) VitisExpDB: a database resource for grape functional genomics. *BMC Plant Biol*, 8, 23.

Duan, K., Ding, X., Zhang, Q., Zhu, H., Pan, A. and Huang, J. (2008) AtCopeg1, the unique gene originated from AtCopia95 retrotransposon family, is sensitive to external hormones and abiotic stresses. *Plant Cell Rep*, 27, 1065-1073.

Elrouby, N. and Bureau, T.E. (2001) A novel hybrid open reading frame formed by multiple cellular gene transductions by a plant long terminal repeat retroelement. *J Biol Chem*, 276, 41963-41968.

English, J.J., Harrison, K. and Jones, J.D.G. (1995) Aberrant transpositions of maize double Ds-like elements usually involve Ds ends on sister chromatids. *Plant-Cell*, 7(8), 1235-1247.

Fengqin, Z., Fangmei, L. and Dabin, G. (1990) Studies on germplasm resources of wild grape species (*Vitis* spp.) in China. In *International 5th Symposium Grape Breeding*. St Martin, Pfalz, Germany: Special issue of VITIS, pp. 50–57.

Feschotte, C., Jiang, N. and Wessler, S.R. (2002a) Plant transposable elements: where genetics meets genomics. *Nature reviews. Genetics*, 3, 329-341.

Feschotte, C., Osterlund, M.T., Peeler, R. and Wessler, S.R. (2005) DNA-binding specificity of rice mariner-like transposases and interactions with Stowaway MITEs. *Nucleic Acids Res*, 33, 2153-2165.

Feschotte, C. and Pritham, E.J. (2006) Mobile DNA: genomes under the influence. *Genome Biol*, 7, 320.

Feschotte, C., Zhang, X. and Wessler, S.R. (2002b) Miniature inverted-repeat transposable elements (MITEs) and their relationship with established DNA transposons. In *Mobile DNA II* (Craig, N.L., Craigie, R., Gallert, M. and Lambowitz, A.M. eds). Washington, DC: ASM Press, pp. 1147-1158.

Finnegan, D.J. (1989) Eukaryotic transposable elements and genome evolution. *Trends in Genetics*, 5, 103-107.

Forneck, A. (2005) Plant Breeding: Clonality — A Concept for Stability and Variability During Vegetative Propagation. In *Progress in Botany*, pp. 164-183.

Franks, T., Botta, R., Thomas, M.R. and Franks, J. (2002) Chimerism in grapevines: implications for cultivar identity, ancestry and genetic improvement. *TAG Theoretical and Applied Genetics*, 104, 192-199.

Galet, P. (1988) *Cépages et vignobles de France* 2nd edn. Montpellier, France: Imprimerie Charles Déhan.

Gerstein, M.B., Bruce, C., Rozowsky, J.S., Zheng, D., Du, J., Korbel, J.O., Emanuelsson, O., Zhang, Z.D., Weissman, S. and Snyder, M. (2007) What is a gene, post-ENCODE? History and updated definition. *Genome Res*, 17, 669-681.

Goremykin, V.V., Salamini, F., Velasco, R. and Viola, R. (2008) mtDNA of Vitis vinifera and the issue of rampant horizontal gene transfer. *Mol Biol Evol*.

Henderson, I.R. and Jacobsen, S.E. (2007) Epigenetic inheritance in plants. *Nature*, 447, 418-424.

Hoen, D.R., Park, K.C., Elrouby, N., Yu, Z., Mohabir, N., Cowan, R.K. and Bureau, T.E. (2006) Transposon-Mediated Expansion and Diversification of a Family of ULP-like Genes. *Mol Biol Evol*, 23, 1254-1268.

Holligan, D., Zhang, X., Jiang, N., Pritham, E.J. and Wessler, S.R. (2006) The Transposable Element Landscape of the Model Legume Lotus japonicus. *Genetics*, 174, 2215-2228.

Huang, J.T. and Dooner, H.K. (2008) Macrotransposition and Other Complex Chromosomal Restructuring in Maize by Closely Linked Transposons in Direct Orientation. *Plant Cell*.

Hudson, M.E., Lisch, D.R. and Quail, P.H. (2003) The FHY3 and FAR1 genes encode transposase-related proteins involved in regulation of gene expression by the phytochrome A-signaling pathway. *Plant J*, 34, 453-471.

Iandolino, A., Nobuta, K., da Silva, F.G., Cook, D.R. and Meyers, B.C. (2008) Comparative expression profiling in grape (Vitis vinifera) berries derived from frequency analysis of ESTs and MPSS signatures. *BMC Plant Biol*, 8, 53.

Imazio, S., Labra, M., Grassi, F., Scienza, A. and Failla, O. (2006) Chloroplast Microsatellites to Investigate the Origin of Grapevine. *Genetic Resources and Crop Evolution*, 53, 1003-1011.

Jaillon, O., Aury, J.M., Noel, B., Policriti, A., Clepet, C., Casagrande, A., Choisne, N., Aubourg, S., Vitulo, N., Jubin, C., Vezzi, A., Legeai, F., Hugueney, P., Dasilva, C., Horner, D., Mica, E., Jublot, D., Poulain, J., Bruyere, C., Billault, A., Segurens, B., Gouyvenoux, M., Ugarte, E., Cattonaro, F., Anthouard, V., Vico, V., Del Fabbro, C., Alaux, M., Di Gaspero, G., Dumas, V., Felice, N., Paillard, S., Jurman, I., Moroldo, M., Scalabrin, S., Canaguier, A., Le Clainche, I., Malacrida, G., Durand, E., Pesole, G., Laucou, V., Chatelet, P., Merdinoglu, D., Delledonne, M., Pezzotti, M., Lecharny, A., Scarpelli, C., Artiguenave, F., Pe, M.E., Valle, G., Morgante, M., Caboche, M., Adam-Blondon, A.F., Weissenbach, J., Quetier, F. and Wincker, P. (2007) The grapevine genome sequence suggests ancestral hexaploidization in major angiosperm phyla. *Nature*, 449, 463-467.

Jiang, N., Bao, Z., Zhang, X., Eddy, S.R. and Wessler, S.R. (2004) Pack-MULE transposable elements mediate gene evolution in plants. *Nature*, 431, 569-573.

Jones, G. and Legge, A. (1987) The grape (*Vitis vinifera* L.) in the neolithic of Britain. *Antiquity* 61, 452-453

Kapitonov, V.V. and Jurka, J. (1999) Molecular paleontology of transposable elements from Arabidopsis thaliana. *Genetica*, 107, 27-37.

Kapitonov, V.V. and Jurka, J. (2001) Rolling-circle transposons in eukaryotes. *Proceedings of the National Acadademy of Sciences of the United States of America*, 98, 8714-8719.

Kapitonov, V.V. and Jurka, J. (2006) Self-synthesizing DNA transposons in eukaryotes. *Proceedings of the National Acadademy of Sciences of the United States of America*, 103, 4540-4545.

Kapitonov, V.V. and Jurka, J. (2008) A universal classification of eukaryotic transposable elements implemented in Repbase. *Nat Rev Genet*, 9, 411-412; author reply 414.

Kawasaki, S. and Nitasaka, E. (2004) Characterization of Tpn1 Family in the Japanese Morning Glory: En/Spm-related Transposable Elements Capturing Host Genes. *Plant Cell Physiol.*, 45, 933-944.

Kidwell, M.G. (2002) Transposable elements and the evolution of genome size in eukaryotes. *Genetica*, 115, 49-63.

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

Le, Q.H., Wright, S., Yu, Z. and Bureau, T. (2000) Transposon diversity in Arabidopsis thaliana. *Proceedings of the National Academy of Sciences*, 97, 7376-7381.

Levadoux, L. (1956) Les populations sauvages et cultivées de Vitis vinifera L. *Ann Amélior Plant*, 6, 59–117.

Liu, D., Bischerour, J., Siddique, A., Buisine, N., Bigot, Y. and Chalmers, R. (2007) The human SETMAR protein preserves most of the activities of the ancestral Hsmar1 transposase. *Mol Cell Biol*, 27, 1125-1132.

Lodhi, M.A., Ye, G.N., Weeden, N.F. and Reisch, B.I. (1994) A simple and efficient method for DNA extraction from grapevine cultivars, *Vitis* species and *Ampelopsis*. *Plant Mol. Biol. Rep.*, 12, 6-13.

Loot, C., Santiago, N., Sanz, A. and Casacuberta, J.M. (2006) The proteins encoded by the pogo-like Lemi1 element bind the TIRs and subterminal repeated motifs of the Arabidopsis Emigrant MITE: consequences for the transposition mechanism of MITEs. *Nucleic Acids Res*, 34, 5238-5246.

Madlung, A. and Comai, L. (2004) The effect of stress on genome regulation and structure. *Annals of botany*, 94, 481-495.

Mao, L., Wood, T.C., Yu, Y., Budiman, M.A., Tomkins, J., Woo, S., Sasinowski, M., Presting, G., Frisch, D., Goff, S., Dean, R.A. and Wing, R.A. (2000) Rice transposable elements: a survey of 73,000 sequence-tagged-connectors. *Genome Research*, 10, 982-990.

Marsoni, M., Bracale, M., Espen, L., Prinsi, B., Negri, A.S. and Vannini, C. (2008) Proteomic analysis of somatic embryogenesis in Vitis vinifera. *Plant Cell Rep*, 27, 347-356.

Matzke, M.A., Mette, M.F. and Matzke, A.J.M. (2000) Transgene silencing by the host genome defense: implications for the evolution of epigenetic control mechanisms in plants and vertebrates. *Plant-Molecular-Biology*, 43(2/3), 401-415.

McClintock, B. (1947) Cytogenetic studies of maize and Neurospora. *Carnegie Inst. Wash. Year Book*, 46, 146-152.

McClintock, B. (1956) Mutations in maize. *Carnegie Inst. Wash. Year Book*, 55, 323-332.

McGovern, P.E. (2003) *Ancient Wine: the Search for the Origins of Viniculture* Princeton, New Jersey: Princeton University Press.

McGovern, P.E., Voigt, M.M., Glusker, D.L. and Exner, L.J. (1986) Neolithic resinated wine. *Nature*, 381, 480-481.

Meyers, B.C., Tingey, S.V. and Morgante, M. (2001) Abundance, distribution, and transcriptional activity of repetitive elements in the maize genome. *Genome Research*, 11, 1660–1676.

Miskey, C., Papp, B., Mates, L., Sinzelle, L., Keller, H., Izsvak, Z. and Ivics, Z. (2007) The ancient mariner sails again: transposition of the human Hsmar1 element by a reconstructed transposase and activities of the SETMAR protein on transposon ends. *Mol Cell Biol*, 27, 4589-4600.

Moreno-Vazquez, S., Ning, J.C. and Meyers, B.C. (2005) hATpin, a family of MITE-like hAT mobile elements conserved in diverse plant species that forms highly stable secondary structures. *Plant Molecular Biology*, 58(6), 869-886.

Moroldo, M., Paillard, S., Marconi, R., Fabrice, L., Canaguier, A., Cruaud, C., De Berardinis, V., Guichard, C., Brunaud, V., Le Clainche, I., Scalabrin, S., Testolin, R., Di Gaspero, G.,

Morgante, M. and Adam-Blondon, A.F. (2008) A physical map of the heterozygous grapevine 'Cabernet Sauvignon' allows mapping candidate genes for disease resistance. *BMC Plant Biol*, 8, 66.

Muehlbauer, G.J., Bhau, B.S., Syed, N.H., Heinen, S., Cho, S., Marshall, D., Pateyron, S., Buisine, N., Chalhoub, B. and Flavell, A.J. (2006) A hAT superfamily transposase recruited by the cereal grass genome. *Mol Genet Genomics*, 275, 553-563.

Naito, K., Cho, E., Yang, G., Campbell, M.A., Yano, K., Okumoto, Y., Tanisaka, T. and Wessler, S.R. (2006) Dramatic amplification of a rice transposable element during recent domestication. *Proc Natl Acad Sci U S A*, 103, 17620-17625.

Nonis, A., Ruperti, B., Pierasco, A., Canaguier, A., Adam-Blondon, A.F., Di Gaspero, G. and Vizzotto, G. (2008) Neutral invertases in grapevine and comparative analysis with Arabidopsis, poplar and rice. *Planta*.

Pejić, I., Maletić, E., Karoglan Kontić, J., Kozina, B. and Mirosević, N. (2000) Diversity of Autochthonous Grapevine Genotypes in Croatia. *Acta Hort. (ISHS)*, 528.

Pelsy, F. and Merdinoglu, D. (2002) Complete sequence of Tvv1, a family of Ty 1 copia-like retrotransposons of Vitis vinifera L., reconstituted by chromosome walking. *TAG, Theoretical and Applied Genetics*, 105, 614-621.

Pindo, M., Vezzulli, S., Coppola, G., Cartwright, D.A., Zharkikh, A., Velasco, R. and Troggio, M. (2008) SNP high-throughput screening in grapevine using the SNPlex genotyping system. *BMC Plant Biol*, 8, 12.

Piriyapongsa, J. and Jordan, I.K. (2007) A Family of Human MicroRNA Genes from Miniature Inverted-Repeat Transposable Elements. *PLoS ONE*, 2, e203. doi:210.1371/journal.pone.0000203.

Rabinowicz, P.D., Palmer, L.E., May, B.P., Hemann, M.T., Lowe, S.W., McCombie, W.R. and Martienssen, R.A. (2003) Genes and transposons are differentially methylated in plants, but not in mammals. *Genome Res*, 13, 2658-2664.

Robertson, H.M. (2002) Evolution of DNA transposons in eukaryotes. In *Mobile DNA II* (Craig, N.L., Craigie, R., Gallert, M. and Lambowitz, A.M. eds). Washington, DC: ASM Press, pp. 1093-1110.

Rogers, D.J. and Rogers, C.F. (1978) Systematics of North American Grape Species. *Am. J. Enol. Vitic.*, 29, 73-78.

Sanmiguel, P. and Bennetzen, J.L. (1998) Evidence that a recent increase in maize genome size was caused by the massive amplification of intergene retrotransposons. *Annals of Botany*, 82(SupA), 37-44.

Schmidt, T. (1999) LINEs, SINEs and repetitive DNA: non-LTR retrotransposons in plant genomes. *Plant-Molecular-Biology*, 40(6), 903-910.

Slotkin, R.K., Freeling, M. and Lisch, D. (2005) Heritable transposon silencing initiated by a naturally occurring transposon inverted duplication. *Nat Genet*, 37, 641-644.

Smalheiser, N.R. and Torvik, V.I. (2005) Mammalian microRNAs derived from genomic repeats. *Trends Genet*, 21, 322-326.

Terral, J.F. (2002) Quantitative anatomical criteria for discriminating wild grapevine (*Vitis vinifera* ssp. *sylvestris*) from cultivated vines (*Vitis vinifera* ssp. *vinifera*). *British Archaeological Reports, International Series*, 1063, 59-64.

This, P., Lacombe, T., Cadle-Davidson, M. and Owens, C.L. (2007) Wine grape (Vitis vinifera L.) color associates with allelic variation in the domestication gene VvmybA1. *Theor Appl Genet*, 114, 723-730.

This, P., Lacombe, T. and Thomas, M.R. (2006) Historical origins and genetic diversity of wine grapes. *Trends Genet*, 22, 511-519.

Thornburg, B.G., Gotea, V. and Makalowski, W. (2006) Transposable elements as a significant source of transcription regulating signals. *Gene*, 365, 104-110.

van Leeuwen, H., Monfort, A. and Puigdomenech, P. (2007) Mutator -like elements identified in melon, Arabidopsis and rice contain ULP1 protease domains. *Molecular Genetics and Genomics*, 277, 357-364.

Vastenhouw, N.L. and Plasterk, R.H.A. (2004) RNAi protects the Caenorhabditis elegans germline against transposition. *Trends-in-Genetics*, 20(7), 314-319.

Velasco, R., Zharkikh, A., Troggio, M., Cartwright, D.A., Cestaro, A., Pruss, D., Pindo, M., Fitzgerald, L.M., Vezzulli, S., Reid, J., Malacarne, G., Iliev, D., Coppola, G., Wardell, B., Micheletti, D., Macalma, T., Facci, M., Mitchell, J.T., Perazzolli, M., Eldredge, G., Gatto, P., Oyzerski, R., Moretto, M., Gutin, N., Stefanini, M., Chen, Y., Segala, C., Davenport, C., Dematte, L., Mraz, A., Battilana, J., Stormo, K., Costa, F., Tao, Q., Si-Ammour, A., Harkins, T., Lackey, A., Perbost, C., Taillon, B., Stella, A., Solovyev, V., Fawcett, J.A., Sterck, L., Vandepoele, K., Grando, S.M., Toppo, S., Moser, C., Lanchbury, J., Bogden, R., Skolnick, M., Sgaramella, V., Bhatnagar, S.K., Fontana, P., Gutin, A., Van de Peer, Y., Salamini, F. and Viola, R. (2007) A high quality draft consensus sequence of the genome of a heterozygous grapevine variety. *PLoS ONE*, 2, e1326.

Verries, C., Bes, C., This, P. and Tesniere, C. (2000) Cloning and characterization of Vine-1, a LTR-retrotransposon-like element in Vitis vinifera L., and other Vitis species. *Genome*, 43(2), 366-376.

Vezzulli, S., Troggio, M., Coppola, G., Jermakow, A., Cartwright, D., Zharkikh, A., Stefanini, M., Grando, M.S., Viola, R., Adam-Blondon, A.F., Thomas, M., This, P. and Velasco, R. (2008) A reference integrated map for cultivated grapevine (Vitis vinifera L.) from three crosses, based on 283 SSR and 501 SNP-based markers. *Theor Appl Genet*, 117, 499-511.

Vicient, C.M., Suoniemi, A., Anamthawat-Jonsson, K., Tanskanen, J., Beharav, A., Nevo, E. and Schulman, A.H. (1999) Retrotransposon BARE-1 and its role in genome evolution in the genus Hordeum. *Plant Cell*, 11(9), 1769-1784.

Weil, C. and Martienssen, R. (2008) Epigenetic interactions between transposons and genes: lessons from plants. *Curr Opin Genet Dev*, 18, 188-192.

Wessler, S.R. (1991) The maize transposable Ds1 element is alternatively spliced from exon sequences. *Mol Cell Biol*, 11, 6192-6196.

Wicker, T., Sabot, F., Hua-Van, A., Bennetzen, J.L., Capy, P., Chalhoub, B., Flavell, A., Leroy, P., Morgante, M., Panaud, O., Paux, E., SanMiguel, P. and Schulman, A.H. (2007) A unified classification system for eukaryotic transposable elements. *Nat Rev Genet*, 8, 973-982.

Xiao, H., Tattersall, E.A., Siddiqua, M.K., Cramer, G.R. and Nassuth, A. (2008) CBF4 is a unique member of the CBF transcription factor family of Vitis vinifera and Vitis riparia. *Plant Cell Environ*, 31, 1-10.

Yu, Z., Wright, S.I. and Bureau, T.E. (2000) Mutator-like Elements in Arabidopsis thaliana: Structure, Diversity and Evolution. *Genetics*, 156, 2019-2031.

Zabala, G. and Vodkin, L. (2007) Novel exon combinations generated by alternative splicing of gene fragments mobilized by a CACTA transposon in Glycine max. *BMC Plant Biology*, 7, 38.

Zabala, G. and Vodkin, L.O. (2005) The wp Mutation of Glycine max Carries a Gene-Fragment-Rich Transposon of the CACTA Superfamily. *Plant Cell*, 17, 2619-2632.

Zohary, D. (1996) The Domestication of the Grapevine *Vitis vinifera* L. in the Near East. In *The Origins and Ancient History of Wine* (McGovern, P.E., Fleming, S.J. and Katz, S.H. eds). New York: Gordon and Breach, pp. 23-30.