A Study of Endogenous Transposon Activity

in Grapevine (*Vitis vinifera* L.*)*

A thesis

submitted in partial fulfilment

of the requirements for the degree of

Doctor of Philosophy

at

Lincoln University

by

Darrell K. Lizamore

Lincoln University

2013

Abstract of a thesis submitted in partial fulfilment of the requirements for the degree of Doctor of Philosophy.

A Study of Endogenous Transposon Activity in Grapevine (*Vitis vinifera* L.)

by

Darrell K. Lizamore

Transposable elements (TEs) are recognised as a significant and ubiquitous component of eukaryotic genomes. This thesis contributes to the current knowledge of these elements by describing the stimulated mobilisation of multiple class I and class II elements from eight TE superfamilies in grapevine somatic embryo cultures following stress treatments and tissue culture, leading to the production of new vegetative material.

An *in silico* analysis of class I TEs in the grapevine genome revealed that although the majority of the 137 defined retrotransposon families exist mainly as eroded fragments, several families show evidence of recent mobility or appear in transcript databases. Based on these results, a high-resolution S-SAP technique was used to identify insertion polymorphisms of three Ty1-*Copia* TE families (*Edel*, *Noble* and *Cremant*) and one Ty3-*Gypsy* family (*Gret1*) across 32 grapevine genotypes, demonstrating the contribution of these elements to genetic diversity in *Vitis*.

By supplementing bacterial suspensions with an organosilicone surfactant, the efficiency of *Agrobacterium*-mediated transient transformation of grapevine leaf tissue was improved by an average of 72-fold. This protocol was used to show that of the above four TE families, only the long-terminal repeat (LTR) sequence of *Edel* is able to drive expression of reporter genes in grapevine leaf tissue, but all four are capable of stimulating expression in the model plant *N. tabaccum*. After two generations, the LTR sequences of *Gret1* and *Edel* no longer induced reporter gene expression in stable *N. tabaccum* transgenic plants, but the LTR sequences of *Cremant* and *Noble* retained a wound-responsive expression pattern.

Due to their sessile lifestyle, plants are forced to endure and adapt to environmental challenges. Biotic (e.g. pathogen attack) and abiotic (e.g. wounding / drought) stress events have previously been shown to stimulate the activity of certain TEs in plants. In this study, transcripts all four TE families (*Gret1*, *Edel*, *Cremant* and *Noble*) were found to increase when Pinot noir embryogenic callus (EC)

cultures were co-cultivated with live yeast species endemic to New Zealand vineyards. Abiotic stresses and fungal extracts did not elicit the same response. A total of 24 new TE polymorphisms, relating to all four of the TE families analysed, were detected by S-SAP in a population of 183 vines regenerated from EC cultures. The majority (14) of these polymorphisms were found in vines regenerated from yeast-stressed EC tissue. The regenerated vines also displayed a variety of phenotypic abnormalities.

Finally, whole-genome sequence data from twenty of the regenerated vines revealed that vines passaged through somatic embryogenesis experienced a general activation of the mobilome, resulting in an average of 64 new TE insertions per plant from both TE classes. Yeast stress at the embryogenic callus stage increased the number of new TE insertions identified by 63%. Despite a strong general bias against coding DNA sequence (CDS) insertions, approximately 2 insertions were found in this context per plant. These data are discussed with regards to the biological implications of endogenous TE mobilisation and the potential use of these elements for saturation mutagenesis in grapevine.

Keywords: genomics, grapevine, mobilome, somatic embryogenesis, somatic mutation, retrotransposon, transposable element, transposon display, *Vitis vinifera*, VvMYBA1, whole-genome sequencing.

Acknowledgements

Research is a long an expensive process, but I have been truly humbled by the number of people who have invested time, energy and finances into this project.

My supervisor, Dr Chris Winefield, has been unfailing in his support and assistance throughout my doctoral studies. I have the greatest admiration for his ability to be at the same time both enthusiastic and critical towards my work and I am fortunate to have gained both a friend and a mentor in these past few years. I am also especially thankful to my co-supervisor, Prof Brian Jordan, for his invaluable advice and feedback at all stages of the project, both in academic and personal regards.

My work has been funded by a Lincoln University Doctoral Scholarship and the inaugural New Zealand Winegrowers (NZW) Rod Bonfiglioli Memorial Scholarship. I am most grateful to these two groups for this opportunity. Specifically, I would like to thank Dr Simon Hooker at NZW for his positivity and support.

Dr Susan Thompson (bioinformatician, Plant & Food Research, New Zealand) has been crucial to the analysis of the raw Illumina sequence data, and Dr Mark Fiers (bioinformatician, VIB, Belgium) has also given much of his time to help us interpret this data. To both of these people I am gratefully indebted. Thanks also to Dr Mat Goddard (Auckland University) for sending us the six yeast strains that his research group has collected from New Zealand vineyards and characterised.

To my lab colleagues, thank you for your friendship and support over the past few years. It has been a pleasure to work with you. Thanks especially to Josh Philips for his management of the laboratory.

During this work a series of earthquakes devastated Christchurch. What followed was the greatest display of community and generosity that I have ever witnessed. To the Cantabrians who have been resilient and made us feel at home throughout this difficult period, we will always speak well of you. Thank you especially to the Metcalf family, who have been our family away from home in every way possible and to David and Eriko Viviers, for their unfailing support, friendship and understanding.

Finally, I thank God for this incredible opportunity and the strength to meet these challenges. To my wife Cathy, who has been kind, patient and selfless, even while working on her own doctoral thesis, I owe my joy and sanity to you.

Table of Contents

List of Tables

List of Figures

Appendices

Acronyms

ABA	abscisic acid
AFLP	amplified fragment length polymorphism
ANOVA	analysis of variance
BAP	6-benzyl amino purine
bp	base pair(s)
CaMV	*cauliflower mosaic virus*
cDNA	complementary DNA
CDS	coding DNA sequence
Cq	quantification cycle
CR	Centromeric retrotransposon
CRE	*cis*-regulatory element
dH_2O	ultrapure water (18.2MW.cm resistivity)
DNA	deoxyribose nucleic acid
dsRNA	double-stranded RNA
EC	embryogenic callus
EDTA	ethylenediaminetetraacetic acid
EMS	ethyl methanesulfonate
EST	expressed sequence tag
gDNA	genomic DNA
GFP	green fluorescent protein
GM	genetically modified
GUS	ß-glucuronidase
HRM	high-resolution melt
IRAP	inter-retrotransposon amplified polymorphism
kb	kilobase pairs
LTR	long terminal repeat
LINE	long interspersed repetitive element
MCS	multiple cloning site
MITE	miniature inverted-repeat transposable element
mRNA	messenger RNA
NGS	next-generation sequencing
NOA	β-naphtoxyacetic acid
NTC	no template control
ORF	open reading frame
PAGE	polyacrylamide gel electrophoresis
PBS	primer binding site
PCR	polymerase chain reaction
PPT	polypurine tract
PR	pathogenesis-related
qPCR	quantitative PCR
RAPD	rapid amplification of polymorphic DNA
RdDM	RNA-dependant DNA methylation
RDR	RNA-dependent RNA polymerase
REMAP	retrotransposon-microsatellite amplified polymorphism

rfu	relative fluorescent units
RISC	RNA-induced silencing complex
RITS	RNA induced transcriptional gene-silencing
RNA	ribose nucleic acid
RNAi	RNA interference
S-SAP	sequence-specific amplified polymorphism
SCAR	sequenced characterized amplified region
SD	standard deviation
SE	somatic embryo *or* standard error
siRNA	small interfering RNA
SNP	single nucleotide polymorphism
SSLP	simple sequence length polymorphism
SSR	simple sequence repeat / microsatellite
ssRNA	single-stranded RNA
T-DNA	transfer DNA
TBE buffer	$tris_{(100mM)}$-$borate_{(90mM)}$-$EDTA_{(1mM)}$ buffer
TE	transposable element
TE buffer	$tris_{(10mM)}$-$EDTA_{(1mM)}$ buffer
TIR	terminal inverted repeat
T_m	melting temperature
tRNA	transfer RNA
TSS	transcription start site
UTR	untranslated region
UV	ultraviolet
WGS	whole-genome sequencing

Chapter 1

Introduction

1.1 The economic importance of viticulture

> *"Quickly, bring me a beaker of wine, so that I may wet my mind and say something clever."*
>
> *- Aristophanes, circa 400 BC*

Fossilised seeds found in Oregon, USA and genetic evidence of polyploidisation events date the appearance of grapevines to approximately 50-60 million yeas ago (CHEN AND MANCHESTER 2007; MALACARNE *et al.* 2012). Today, the wild grapevine, *V. vinifera* ssp. *sylvestris*, is rare and sparsely distributed from the Atlantic coast of Europe to the Western Himalayas (GRASSI *et al.* 2008). Habitat disruption and susceptibility to New World pathogens such as *Phylloxera* have led to a severe reduction in distribution. In 2000 wild grapevine was recognised as critically endangered and as a result is now protected in several European states (GRASSI *et al.* 2006).

In contrast, through a particularly successful history of agriculture and trade spanning the past six millennia, the modern Eurasian grapevine (*Vitis vinifera* ssp. *sativa*) has risen to become the world's most abundant and economically important fruit crop (www.fao.org). The separation of these two lineages is believed to be the earliest example of the domestication of a fruit-bearing plant.

Genetic analysis of modern grape varieties indicates that domestication occurred on at least two occasions (ARROYO-GARCÍA *et al.* 2002; GRASSI *et al.* 2003), at a similar time and place to the invention of wine (MCGOVERN 2003). The earliest evidence of viniculture was revealed by the chemical analysis of Neolithic pottery uncovered at Hajji Firuz in the northern Zagros Mountains of Iran. Fragments of jars were found to have traces of tartaric acid, a compound only naturally abundant in grapes. Also present on the shards was *Pistacia* tree resin, which is soluble in ethanol and would have been added as a preservative, precluding that the remains are from accidentally fermented grape juice. The jars were dated to approximately 5400 BC (MCGOVERN *et al.* 1996).

With time, grapevine cultivation spread across the Middle East and to nearby Egypt (MCGOVERN *et al.* 2009; BARNARD *et al.* 2011). Ancient civilizations including the Assyrians, Phoenicians, Greeks and Romans planted vineyards first around the temperate Mediterranean and then across Europe, following the major trade routes (THIS *et al.* 2006). Trade between Europe and Asia via the Silk Road brought viticulture to China in approximately 300 BC, where vines were initially planted in the high Turpan Basin and later grown near the imperial palace (JIANG *et al.* 2009). Missionaries carried

grapevine seeds to the Americas in the 16th century and immigrants took cuttings to South Africa and Australia in the 19th century, accomplishing the grapevine's spread to all six inhabited continents (THIS *et al.* 2006).

At the end of the nineteenth century Italian-born viticulturist Romeo Bregato was commissioned by the New Zealand Department of Agriculture to assess the potential for domestic wine production (BRAGATO 1895). The recommendations of his report were used to build an industry that now spans the country, with a current annual production of 178.9 million litres, valued at USD 950 million (2012 data; www.nzwine.com). Although current industry trends show a decrease in European production, New World production continues to show rapid growth, with the 2012 national yield a record in both New Zealand and China (www.oiv.int).

1.2 The value of cultivar identity

At the point of sale, consumers are not usually able to directly determine the quality of a wine before purchase (i.e. by tasting). Therefore, purchase decisions are based on indirect indicators of wine quality. In European markets, the region of origin is typically found to be among the most significant indicators of wine quality, while in New World markets grape variety frequently scores highly (ARIAS-BOLZMANN *et al.* 2003; SCHAMEL AND ANDERSON 2003; MCCUTCHEON *et al.* 2009). The distinction is supposed to be a legacy of European appellations known for production of wines of a certain style. In contrast, the marketing strategies of emerging wine industries have primarily focused on variety and brand.

The grape variety used to produce a given wine is therefore one of the premier labels by which wine producers communicate product quality to potential buyers. In addition to maintain the market value of familiar varieties, clonal propagation of vine material ensures predictability and consistency in plant growth and development. It is unsurprising therefore that the vast majority of currently cultured grapevines represent only a few of the approximately eight thousand varieties known to exist (ALLEWELDT AND POSSINGHAM 1988). The value of clonal propagation of vines has been so established for so long that the crosses that produced the popular Cabernet Sauvignon and Chardonnay varieties are believed to have occurred during the medieval period (BOWERS AND MEREDITH 1997; BOWERS *et al.* 1999).

1.3 Limited diversity within varieties

As an entire species, modern grapevine is comparable in diversity to crops such as maize (MYLES *et al.* 2011). However, due to a uniquely long history of vegetative propagation, the genetic diversity within the relatively few varieties utilised for commercial wine production is very limited. While intervarietal breeding programs can therefore be expected to produce new types

with interesting and desirable characteristics, any progeny will lack the historical pedigree, and therefore the brand recognition, of the famous varieties. Despite the thousands of registered grapevine varieties that currently exist, only a select few are grown for wine production, evidence that the wine industry is not lacking for new varieties. Any genetic improvement must therefore come from within the established popular varieties to be considered for industry acceptance.

1.4 Exploiting natural genetic variation

After a visit to the Sancerre region of France in 2008, New Zealand-based viticulturists Rod Bonfiglioli & Nick Hoskins commented on the advantages that continuous evaluation of new clonal material provides to the French wine industry (BONFIGLIOLI AND HOSKINS 2008). They went on to point out the noticeable absence of such strategies in New Zealand.

The Bonfiglioli & Hoskins report draws attention to a history of successful genetic improvement by certain members of the international wine industry in the form of novel clones of existing varieties. Such clones are the result of DNA mutations occurring in the growth apices of a vine. As the plant grows, portions of the vine containing the new mutation may appear atypical to be of the parental clone. When identified, this material can be collected and propagated by shoot cuttings. With very few exceptions, identification of these 'bud sports' in vineyards and nurseries has been the sole source of new clones for the wine industry. The extensive use of bud sport-derived clones in commercial settings is evidence of the value of this source of genetic variation.

Centuries ago, a mutant form of a single grapevine gene produced a sport incapable of producing anthocyanin, the purple-red pigment typical of ripe wild grape berries (KOBAYASHI *et al.* 2004). Consequently, the ripe berries of such a vine can be used to make white wine. This specific mutation has to date been found in all white grape varieties tested (WALKER *et al.* 2007). The proliferation of this ancestral allele is an example of the profound effect that selective bud-sport propagation has had on our modern conceptions of wine styles.

1.5 Tracking and increasing mutation rates

In contrast with breeding crosses, which shuffle the hereditary material of two parent organisms, bud sport mutations entail a specific change, generally leaving the remainder of the genome unaltered. A vine produced from a cross between Pinot noir and Sauvignon blanc, for example, will exhibit an amalgam of traits inherited from both parental types. A bud sport, however, may differ from the parental vine in only a single trait, but be otherwise identical. This

offers the potential for genetic improvement while maintaining varietal identity, such as a Sauvignon blanc vine with reduced growth vigour.

Unfortunately, natural rates of somatic mutation are difficult to track and identify, particularly when the resulting changes are subtle or not apparent. Vegetative material that acquires a mutation conferring resistance to a pathogen or abiotic stress to which the vine is not exposed in the same season will likely be pruned off and lost.

Therefore, techniques to accelerate and track the rate of bud sport mutations in grapevine are essential for the production of a resource of new clonal material. A collection of clonally propagated grapevine plants possessing a variety of new somatic mutations would provide a valuable resource to fast-track genetic improvement initiatives and addresses the shortage of genetic variation available to New Zealand winegrowers.

1.6 Hypotheses to be tested

The ideas pursued in the course of this research project can be expressed as the following distinct hypotheses:

1.6.1 H_1: The grapevine genome reveals differences among the historical activity of transposon families

Large proportions of eukaryotic genomes have been found to be comprised of repetitive DNA, the majority of which is derived from endogenous mobile DNA sequences known as transposable elements (discussed in section 2.2). This is true of the grapevine sequence published in 2007 (JAILLON *et al.* 2007; VELASCO *et al.* 2007). The public availability of whole-genome draft sequence assemblies and computational tools for searching these data enables *in silico* approach to characterise the integrity and sequence similarity of elements from defined TE families.

The corresponding null hypothesis is:

$H_{0\text{-}1}$: No evidence can be found for differences in the historical activity of transposon families within the grapevine genome sequence data.

1.6.2 H_2: Transposon insertion polymorphisms contribute to clonal variation in grapevines

Anecdotal accounts of high phenotypic variance within clonal vineyards are common among winegrowers, particularly within certain varieties. While some phenotypic changes have been attributed to rearrangements in cell layer organisation, numerous examples have been traced to

mutations caused by TEs (KOBAYASHI *et al.* 2004; FUJINO *et al.* 2005; MOON *et al.* 2006; FERNANDEZ *et al.* 2010; FERNANDEZ *et al.* 2013).

The testable null hypothesis is:

$H_{0\text{-}2}$: Phenotypically different vines grown from vegetative material have identical TE insertion patterns.

1.6.3 H_3: The insertion patterns of specific TE families can be used to track genetic variation in grapevine

Known transposon sequences can be used as tags for insertion mutations. Such a tagged mutagenesis approach simplifies genotyping and genomic characterisation of mutants. In order to demonstrate this, it is necessary to determine the genomic loci of individual TE polymorphisms.

The null hypothesis is:

$H_{0\text{-}3}$: Differences between TE insertion profiles cannot be associated with the presence or absence of a single TE insertion at a specific locus in the genome.

1.6.4 H_4: TEs in grapevines are activated under stress conditions

Most transposable elements contain nucleotide sequences that are common to the regulatory regions of stress response genes. The hypothesis is therefore put forward that exposing plant tissue to certain stress conditions will increase the relative abundance of transcripts from these elements. The null hypothesis to be tested in this case can be stated as:

$H_{0\text{-}4a}$: Transcript levels of grapevine LTR-retrotransposons are not altered by environmental stress events.

As the nucleotide sequences, and therefore the *cis*-regulatory elements present, differ between the various retrotransposon families, the response to the stress events may not be identical among families.

Transcription of TE insertions may trigger epigenetic changes to the genome. However, reverse transcription of the RNA and successful insertion at a new locus is necessary for novel genetic mutations to occur. Confirmation that this complete process has occurred would involve disproving the following null hypothesis:

$H_{0\text{-}4b}$: Multiple grapevine plants regenerated from the same stressed tissue have identical TE insertion patterns.

1.6.5 H_5: An *in silico* analysis of TE-related sequences in the genome enables prediction of the most mutagenic element families

Computational approaches are used to determine retrotransposon targets for marker and transcript analysis, based on evidence of recent mobilisation. However, whole-genome sequencing allows an unbiased analysis of TE mobilisation in mutant plants, allowing verification of whether those TE families for which evidence of recent activity is found are in fact accountable for the highest number of novel mutations.

The null hypothesis is:

$H_{0\text{-}5}$: The current mutagenicity of grapevine TE families cannot be linked to predictors of recent historical TE activity within the grapevine genome sequence data.

1.6.6 H_6: Mobilisation of endogenous retrotransposons can be used to produce phenotypically varied grapevine plants.

Genetic variation introduced by transposon mobilisation can be detected using molecular techniques at the earliest stages of plant development. Mutations of economic interest to the grape industry, however, are those that contribute to altered plant growth and development. These can only be confirmed as the plants mature.

The applicable null hypothesis in this case can be stated as:

$H_{0\text{-}6}$: It is not possible to identify distinct phenotypic abnormalities in a population of grapevines regenerated after transposon mobilisation.

Chapter 2

Literature Review

2.1 Mutations in vegetative tissue result in chimeras

2.1.1 Sexual reproduction in agriculture

The production of gamete cells involves a special type of cell division (meiosis) in which the genetic material is shuffled and reduced by half. The full genetic complement is then restored by the conjugation of gametes from two parents. Due to the scope of possible permutations during meiotic division of the genome, offspring produced in this way are genetically distinct from one another and both parents. The vast majority of genetic variation between individuals within most populations is therefore a direct result of sexual reproduction.

In agriculture, propagation via sexual recombination is frequently used for grain, vegetable and oriental crops. Thousands of seeds can be collected from the best plants and stored for future plantings. This allows a rapid increase in crop size within a few generations. Seed is an ideal tissue type for plant propagation as it is specifically adapted to protect hereditary material for long periods of time, and germinate when conditions are suitable.

Sexually propagated crops may have undergone thousands of rounds of reproduction in the history of their domestication (ZOHARY 2004). At each stage, the population is subject to what Charles Darwin termed the "unconscious selection" of the environment (DARWIN 1870) as well as a conscious selection by the collector of the seed. Consequently, such crops have continuously evolved since their initial domestication and usually appear very different to their wild relatives.

2.1.2 Vegetatively propagated crops

Sexual propagation has, however, multiple disadvantages for agriculture. Although recombination events improve the efficiency of selection by removing negative correlations between favourable alleles (BARTON AND CHARLESWORTH 1998), they also break down linkage between co-adapted genes (RICE 2002). In addition, the genomes of certain crops, such as fruit trees, are highly heterozygous (SCALABRIN *et al.* 2010). Seeds collected from crosses of such crops will mostly produce progeny that have no economical value.

Many plant species, such as grapevine, are capable of the asexual reproduction of whole new individuals from a limited amount of meristematic tissue. Growers have taken advantage of this fact to select the best individuals from a population and clonally propagate that type. Through the

continuous collection and planting of cuttings, farmers can fill their lands with genetically identical plants. Without the variation introduced by sexual crosses, the genetic material remains far more consistent between individuals and generations. The stability of plant growth and development is desirable for the optimisation of farming practices and for predictable crop yield and quality. It has been argued that the adoption of vegetative propagation techniques was an essential step in the development of successful fruit tree agriculture (ZOHARY AND SPIEGEL-ROY 1975).

2.1.3 Somatic mutations

During the lifetime of an organism, a precise copy of the full genome needs to be made at every cell division. Despite the high fidelity of the DNA polymerase enzymes responsible for this task – single base error rates are just 10^{-5} to 10^{-6} (BRITT 1999) – errors are nevertheless inevitable at the scale of whole genomes. Additionally, DNA suffers damage from environmental and chemical factors such as ultraviolet (UV) radiation and free radicals produced by cellular metabolism.

Due to their sessile nature and to facilitate photosynthetic production of organic compounds, autotrophic plants endure a particularly high exposure to solar radiation. The ozone and oxygen in the Earth's atmosphere prevents penetration of high-energy UV-C radiation. However, UV-B radiation reaches the biosphere and is able to penetrate organic tissue. Absorption of UV-B by DNA results in lesions and dimer production between adjacent pyrimidine residues, the most common of which are cyclobutane pyrimidine dimers (CPDs) (JORDAN 2002).

Incorrect DNA repair can result in a variety of different mutation types. Base mismatches produce single nucleotide polymorphisms (SNPs) which can alter the sequence of genes, cause polypeptide strands to stop prematurely and interrupt sequence-specific DNA-protein interactions. Large scale genome rearrangements can arise from sequence insertions, deletions, and cross-linking between strands (BRAY AND WEST 2005). Without correction, mutations will be propagated through subsequent cellular replication cycles, becoming stable.

Since the germ-line tissue comprises a relatively small proportion of an organism, most mutations will occur in somatic tissue. For this reason somatic mutations do not generally affect sexually propagated offspring, and therefore do not normally contribute to the gene pool. However, the propagation of plants by cuttings stabilises somatic mutations in future generations. Somatic mutations therefore have significant impact in vegetatively propagated plant populations.

2.1.4 Chimerism of vegetative propagules

The meristem of most dicotyledonous plants is divided into 2-3 layers, which are separately responsible for the production of different tissues. The epidermal tissue is produced from L1 and

internal tissues from L2 and L3. Grapevine meristem consists of two layers, which can be genetically different (THOMPSON AND OLMO 1963). Mutations that occur in one cell layer of the meristem therefore only persist in tissue produced from this layer (FERNANDEZ *et al.* 2006). Seed are produced exclusively from the L2 layer, whereas vegetative propagation preserves the separation of the cell layers (BERTSCH *et al.* 2005). As somatic mutations accumulate over time, the two layers of the vegetative tissue become more diverse (FRANKS *et al.* 2002; RIAZ *et al.* 2002; HOCQUIGNY *et al.* 2004).

The stable existence of two genetically different cell layers in a plant is known as periclinal chimerism, a state which appears to be common among elite grapevine varieties (RIAZ *et al.* 2002; PELSY *et al.* 2010). Periclinal chimerism is important to the phenotype of several important varieties. The red-grey colour of the berries of Pinot gris, for example, is the result of a mutation preventing pigmentation in only the inner cell layer of the berry skin, produced from L2 meristem. The mutation is not present in the L1 meristem, which forms the pigmented outer layer of the berry skin. All progeny of self-crossed Pinot gris plants produce only white fruit, while plants regenerated from the L1 cell layer produced black fruit (PELSY 2010).

Chimeric genotypes have been separated by passage of the clone through somatic embryogenesis. Franks and colleagues (2002) regenerated vines from somatic embryos of the Champagne variety Pinot Meunier, a chimeric bud sport of Pinot noir. Regenerants were genotyped using a microsatellite marker for which the chimeric Pinot Meunier produces three alleles. All regenerated plants could be separated into two distinct genotypes, each with only two alleles at the tested locus. Plants of one group were both genetically and phenotypically indistinguishable from Pinot noir, while those of the other genotype displayed a novel dwarf phenotype with short internode lengths and hairy leaves.

2.1.5 Grapevine genetic improvement strategies

As the wine industry continues to grow and be challenged by the demands of new markets, major goals for genetic improvement remain (VIVIER AND PRETORIUS 2002). Viticulture has a long history of the use of somatic mutations for the step-wise development of elite clones within important varieties. Nevertheless, this approach also has serious disadvantages. Selecting plants based purely on phenotype means that subtle and invisible changes are likely to never be identified. Mutants are likely to appear as chimeras, affecting only a fraction the plant tissue. Additionally, the low natural rates of somatic mutation mean that interesting sports are usually chance identifications by viticulturists.

Breeding strategies

Breeding programs are a staple of crop improvement strategies (for review see (VARSHNEY *et al.* 2005; MOOSE AND MUMM 2008; TESTER AND LANGRIDGE 2010) and can be used to exploit the diversity between grapevine cultivars. However, they require large populations and multiple generations to overcome the random linkage disequilibrium between loci in the genome (FELSENSTEIN 1974). Grapevine breeding programs are particularly slow and expensive, due the long generation times, annual reproductive cycle and vineyard management and space requirements of the crop (CHAÏB *et al.* 2010). Finally, as mentioned (see 2.1.2) progeny produced from crosses between varieties lose the value of varietal identity, and show very high variation due to the heterozygosity of the parents.

Transgenesis

The value of introducing or silencing specific genes in an otherwise stable genetic background has led transgenesis to become an essential tool in the field of functional genomics. Successful use of transgenesis for the improvement of various crop species has resulted in sustained optimism towards the genetic improvement of grapevine (ALLEWELDT AND POSSINGHAM 1988; VIVIER AND PRETORIUS 2002; VIDAL *et al.* 2010).

Although simple and elegant as a theoretical technique, grapevine transformation has multiple technological limitations. No technique has yet been reported for the initiation and maintenance of appropriate explant cultures from all grapevine varieties. Additionally, transformation experiments generally report low efficiencies and can be complicated by chimerism of regenerated plants. These factors, together with the long generation times required for full characterisation of transgenic vine phenotypes, have made the technique slow and problematic in practice.

Furthermore, environmental and ethical concerns have restricted the uptake of genetically modified (GM) crops for agriculture. The first field trials of transgenic vines were set up in 1996 in the Champagne vineyards of Moêt & Chandon by the French Institut National de la Recherche Agronomique (INRA). The project showed that certain transgenic rootstock lines expressing the *grapevine fanleaf virus* (GFLV) coat protein gene proved resistant to GFLV infection (VIGNE *et al.* 2004). However, increased public antagonism towards GM crops, and opposition to this study in particular, resulted in the early termination of the field trials.

Nevertheless, INRA decided to proceed with further expanded field trials of transgenic rootstocks. Prior to initiation, a local monitoring committee comprised of representatives of the research team, local public councils and the winegrowing profession assessed the trial protocol. Since the local winegrowing profession region refused to allow GM plants to be trialled within the perimeter of Alsace vineyards, the project had to be conducted at the INRA centre in Colmar. Two hundred public meetings and conferences were held in order to discuss the project. But after the project was

eventually initiated, the vines were destroyed by activists on two occasions, and INRA was finally forced to abandon the research (LEMAIRE *et al.* 2010).

Despite widespread public and industry controversy regarding GM vines, the technology has had some traction. In the period up until 2010, 66 transgenic grapevine field trials were conducted worldwide, but only seven of these were performed in Europe (www.gmo-compass.org).

Saturation Mutagenesis

While transgenesis involves the production of a few plant lines with very specific genomic changes, an alternative approach is to create large plant populations with random mutations. With the production of sufficiently large populations, an individual harbouring a mutation in almost any gene can expect to be found. This technique, known as saturation mutagenesis, has been used been used with great success both in functional genomics studies and for crop improvement. Chemical mutagens, most commonly ethyl methanesulfonate (EMS) (KOORNNEEFF *et al.* 1982), and fast neutron irradiation (LI *et al.* 2001) have both been successfully used in plant systems. Radiation mutagenesis experiments were particularly popular in the 1950s as part of the "Atoms for Peace" program in the United States, which sought to establish peaceful uses for atomic energy. But as the political incentive to use radiation has decreased EMS mutagenesis has become more common. However, the random nature of these types of mutagenesis means that the identification of mutation alleles is not a trivial task.

The successful use of EMS and radiation mutagenesis has been well-established, contributing to the understanding of such crucial plant systems as flowering (MICHAELS AND AMASINO 1999), herbicide resistance (JANDER *et al.* 2003) and UV-B response (FAVORY *et al.* 2009). The advantages of SNP mutagenesis lie in the broad applicability of the techniques to any species, the improved identification of mutant loci by the addition of the TILLING technique (MCCALLUM *et al.* 2000) and an abundance of genotyping approaches (KWOK 2001).

Since the mid 1980s, the pathogenic bacterium *Agrobacterium tumefaciens* has been co-opted for the transformation of a variety of cultivated plant species (HORSCH *et al.* 1985; GELVIN 2003). During infection of plant host cells, *A. tumefaciens* transfers a small section of DNA, known as transfer DNA (T-DNA) into the host nucleus to be incorporated into the host genome. Insertion of the T-DNA causes disruption of the sequence at random sites in the host genome. The process of T-DNA insertion can therefore be used as a mutagenic tool. While large sequence insertions are more likely to have an impact on the genome than SNPs caused by radiation, the real advantage of this technique is that the foreign T-DNA sequence introduces a unique tag by which mutation sites can be mapped. Tagged mutagenesis greatly facilitates genotyping of mutant progeny and therefore the association of genotype with phenotype. Insertional mutagenesis with T-DNA has been used for

large-scale saturation mutagenic studies, most notably in *Arabidopsis thaliana* (ALONSO *et al.* 2003) and rice (JEON *et al.* 2000).

Recently, transposable elements (TEs) have been used for saturation mutagenesis. These repetitive genetic sequences are defined by their capacity to mobilise and produce mutations within a genome. Although most of TEs are silent within their host genome, their activity is often not supressed when they are transplanted into the genome of a different species. Moreover, the mobility of certain non-autonomous elements depends on the expression of genes from autonomous TEs. Co-transformation with a non-autonomous TE and the auxiliary genes of an autonomous element can be used generate mutations in a target genome. The auxiliary genes can then back-crossed out to halt the transposon's mobility, while preserving any novel mutations (LISCH 2002). TEs have recently been used to develop mutagenized populations of several model plant species (described later in 2.2.8).

2.2 Transposable Elements

> *"Only a decade ago, most scientists thought humans had about 100,000 genes"*
>
> *- Francis S. Collins, Oct 2004*
> *(Director, National Human Genome Research Institute)*

With over 180 eukaryotic genomes now fully sequenced, the sparseness of genes found per genome has come as a surprise to the scientific community. Before the human genome was fully sequenced, most professional estimates were that there existed in the order of 100,000 to 150,000 genes. The current count of known protein-coding genes stands at 20,769 (www.ensembl.org/Homo_sapiens, retrieved 18th Oct 2013), the exons of which represent approximately 1% of the total genome sequence (COFFEY *et al.* 2011). However, since having been famously labelled "Junk DNA" four decades ago (OHNO 1972), an appreciation for the importance of non-coding regions of the genome has progressively increased.

In 1948, Barbara McClintock described the activities of two unusual loci in the maize genome. Following a series of breeding experiments, she reported the instability of genes that produce pigmentation in the aleurone layer of kernels. She noted that frequent chromosome breaks at the Dissociation (*Ds*) locus on the short arm of chromosome 9 were responsible for chimeric patches of colour loss in maize kernels. Furthermore, she showed that sequence changes at the Ds locus were not only heritable, but reversible, and under the control of a second 'controlling element', which she named Activator (*Ac*) (MCCLINTOCK 1948; MCCLINTOCK 1951). Though she had been the first to produce a genetic map of maize, she realised when trying to map the location of these new loci that they could not consistently be located on the same chromosome (MCCLINTOCK 1950). Despite having received recognition for her previous cytogenetic research, McClintock struggled to convince her peers of the activity of controlling elements and consequently stopped publishing her data.

As recent technologies have rapidly accelerated the mapping and sequencing of whole genomes, McClintock's unstable genetic sequences have been proved ubiquitous, accounting for large proportions of non-coding DNA in eukaryotes. They are now known as transposable elements (TEs) or transposons, for their characteristic ability to change location within the genome, a process called transposition. A strong correlation exists between the proportion of transposable elements within a genome and the size of that genome (r=0.99; (TENAILLON *et al.* 2010). *Arabidopsis thaliana* cv. *Columbia*, with a genome of approximately 125Mb, for example, consists of approximately 10% TEs (ARABIDOPSIS GENOME INITIATIVE 2000), whereas over 85% of the 2.3Gb maize genome is made up of TEs (SCHNABLE *et al.* 2009).

In 1983, McClintock's work regarding the discovery of TEs was finally recognised with the presentation of that year's Nobel Prize for medicine, the only occasion that a woman has won the prize unshared.

2.2.1 Transposon types

Transposable elements are classified according to the presence and arrangement of their functional sequences (WICKER *et al.* 2007).

Class I: Retrotransposons

The class I elements (retrotransposons) consist of those elements whose mobilisation involves the production of an RNA intermediate, which is then reverse-transcribed to produce a new DNA copy (Figure 2.1 A). This "copy-and-paste" process results in an increase in the size of the host genome. The class includes elements that are flanked by identical long terminal repeats (LTR retrotransposons) as well as non-LTR retrotransposons. The non-LTR retrotransposons are further classed as either long interspersed repetitive elements (LINEs) or short interspersed repetitive elements (SINEs).

The LINEs are similar in structure to the internal portion of LTR-retrotransposons, with the same basic gene arrangement, but lacking flanking repeat sequences. Their sequence diversity suggests that LINEs may have been ancestral to the LTR-retrotransposons (XIONG AND EICKBUSH 1990). The much smaller SINEs are descendant of RNA polymerase II products, such as transfer RNA (tRNA), which have acquired the ability to be reverse transcribed and re-integrated into the genome by the gene products of LTR-retrotransposons and LINEs (SCHMID 1998).

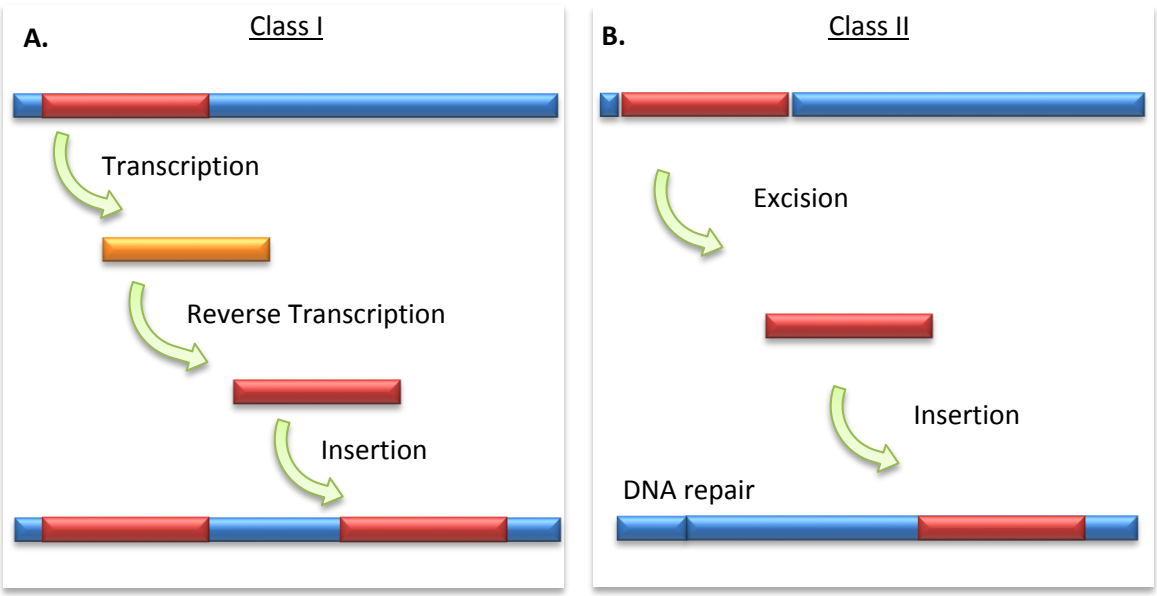


Figure 2.1 Schematic representation of transposon mobility. (A) Transposition of a class I element via an RNA intermediate. (B) Transposition of a class II element via a DNA intermediate.

Class II: DNA transposons

The characteristic structure of a DNA transposon includes a single open reading frame, flanked by terminal inverted repeats (TIRs). Unlike retrotransposons, DNA transposon mobility generally involves the complete excision of the element, followed by reinsertion elsewhere in the genome (Figure 2.1 B). A single transposase gene controls the autonomous mobility of class II TEs. No RNA intermediate is produced and this "cut-and-paste" mobilisation does not generally increase the genome size (KUNZE *et al.* 1997). The exceptions to this rule are the *Helitron* and *Maverick* superfamilies of Class II TEs, whose unusual mobilisation does result in a copy number increase (KAPITONOV AND JURKA 2007; PRITHAM *et al.* 2007).

The class number of a given element type is therefore indicative of the number of DNA strands that are cut at the donor site during transposition. Ten to seventeen defined superfamilies of class II transposon have been described, of which five (*CACTA*, *hAT*, *Mutator*, *PIF/Harbinger*, and Tc1/*Mariner*) have been found in plant genomes (FESCHOTTE AND PRITHAM 2007; YUAN AND WESSLER 2011).

Apart from full-length DNA transposons, miniature inverted-repeat transposable elements (MITEs) are also present in eukaryotic genomes. These short, highly repetitive sequences are characterised by TIRs without internal gene-coding sequences. They are believed to derive from class II elements that have suffered internal deletions, and are therefore incapable of autonomous transposition (FESCHOTTE AND MOUCHES 2000).

Both classes of TE contain autonomous and non-autonomous elements. Despite the loss of functional genes, non-autonomous TEs are not 'dead'. Instead, their mobilisation is dependant on the availability of proteins produced by autonomous TEs (KUNZE *et al.* 1997; KUMAR AND BENNETZEN 1999) .

2.2.2 A closer look at LTR-retrotransposons

Retrotransposons are the most abundant repeat type in eukaryotic genomes. Their high copy number is due to the nature of their mobilisation, which produces new insertions while preserving the original element.

Structure

In structure, retrotransposons are highly similar to the retroviruses that infect vertebrate genomes. The distinguishing feature of retroviruses is the presence of an *env* gene, which encodes transmembrane proteins required for intercellular transfer of virions, enabling infection. Retrotransposons have been phylogenetically separated into two superfamilies, Ty1-*copia* and Ty3-*gypsy*, named for the first elements of each group to be identified. Elements from both groups are flanked by identical LTRs, which can be divided into unique 3' (U3), repeated (R) and unique 5' (U5) regions. The groups differ, however, in the arrangement of the domains within their internal genes (Figure 2.2).

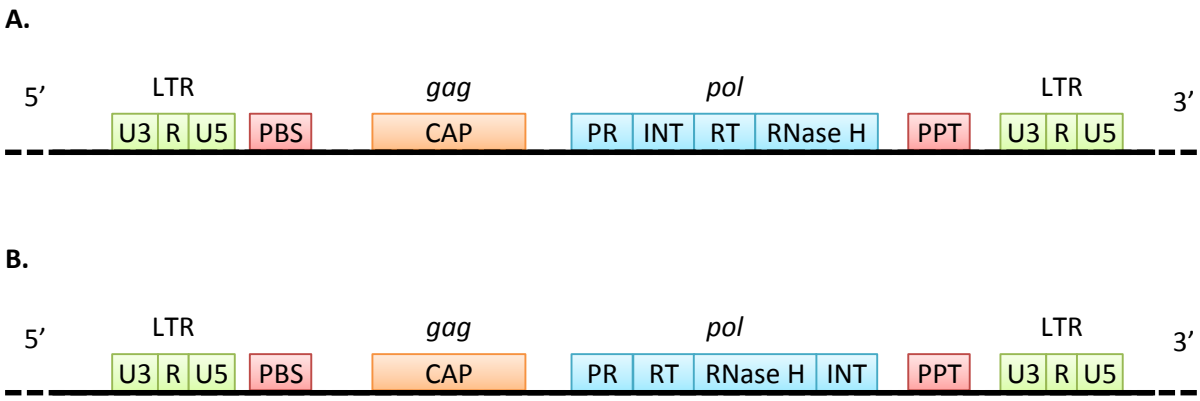


Figure 2.2 Arrangement of structural components of Ty1-*Copia* (A) and Ty3-*Gypsy* (B) LTR-retrotransposons within genomic DNA. Note the differing domain orders within the *pol* gene. Key sequences include: U5 (unique 5' region), R (repeated region), (U3) unique 3' region, PBS (primer binding site), CAP (capsid-like-proteins), PR (protease), INT (integrase), RT (reverse transcriptase), PPT (polypurine tract) modified from the work of Kumar and Bennetzen (1999). Diagram is not drawn to scale.

The coding region of a retrotransposon is made up of two genes: *gag*, which produces proteins responsible for processing and packaging of the transposon RNA, and a multi-domain *pol* gene. The genic sequences are translated as a single polyprotein, which is cleaved into individual functional proteins through the action of the protease domain within the *pol* gene. The *pol* gene also contains

domains for reverse transcriptase and RNase H proteins responsible for the replication of the retrotransposon. Finally, an integrase enzyme, encoded by the INT domain of the *pol* gene, catalyses the insertion of the new element copy into the genomic DNA (KUMAR AND BENNETZEN 1999).

In contrast to vertebrates, plants are not known to suffer from deadly diseases caused by retroviruses. However, examples of plant retrotransposons have been identified which contain genes whose predicted protein structure is similar to that of *env* gene products (PETERSON-BURCH *et al.* 2000; VICIENT *et al.* 2001). Such elements have frequently been referred to as plant endogenous retroviruses, although examples of intercellular transport or infection have not yet been reported. Certain TE families do show high sequence similarity across plant species boundaries, however, suggesting that horizontal transfer of these elements between plant species may have occurred in the past (ROULIN *et al.* 2009; DU *et al.* 2010).

While the sequence similarity of retroviruses and retroelements has convinced most researchers of their common ancestry, the question of chronology is less clear. Examples exist of retrotransposon families that appear to have acquired an *env* gene (HANSEN AND HESLOP-HARRISON 2004), as well as endogenous retroviruses which seem to have lost this gene (YANO *et al.* 2005).

Taxonomic divisions

Elements of the Ty1-*Copia* and Ty3-*Gypsy* superfamilies have been divided into two major branches each and further subdivided into named clades according to the sequence similarity of the *pol* gene (LLORENS *et al.* 2009). Only elements from branch 2 of Ty1-*Copia* have been identified in plants, whereas elements from both branches of Ty3-*Gypsy* have been found.

Although a large number of elements have been identified, repetitive elements tend to be the most difficult sequence data to reassemble from short reads. Consequently, mobilome data are commonly incomplete and frequently constitute much of the unassembled portions of genome projects. The division of the retrotransposon superfamilies into distinct clades is therefore still a work in progress. Table 2.1 presents the major retrovirus clades existent in plants, as described by three recent publications.

Table 2.1 Taxonomic division of retrotransposon superfamilies in plants, according to three recent publications.

(HŘIBOVÁ *et al.* 2010)	(LLORENS *et al.* 2011)	(DOMINGUES *et al.* 2012)
Ty1-*Copia*		
Branch 2		
Angela	Tork	Angela
TONT1		Tar
TNT1		
SIRE1 / Maximus	Sire	Maximus / Sire
	Oryco	Ivana / Oryco
Hopscotch	Retrofit	Ale / Retrofit
		Bianca
Tos17		
Ty3-*Gypsy*		
Branch 1 (Chromoviruses)		
CRM	CRM	CRM
Tekay	Del	Del / Tekay
Galadriel	Galadriel	Galadriel
Reina	Reina	Reina
Branch 2		
Athelia	Athelia	Athelia / Tat
Tat	Tat	
Ogre		

Mobilisation

Transcription of LTR-retrotransposons begins at the R region of the 5' LTR and proceeds through the R region of the 3' LTR (Figure 2.3). The primer binding side (PBS) just inside the 5' LTR is homologous to an endogenous RNA, generally $tRNA^{MET}$ in plants. Binding of the tRNA primes reverse transcription by the reverse transcriptase enzyme released by cleavage of the retrotransposon-derived *pol* polypeptide. Reverse transcription proceeds in an upstream direction to produce a short cDNA strand complementary to the 5' end of the retrotransposon RNA. The RNase H from the cleaved *pol* polypeptide digests RNA in the hybrid portion of the molecule, releasing the cDNA. The single-stranded DNA has homology to the R region of the 3' LTR, where it primes reverse transcription of the remainder of the first cDNA strand. After RNase H again digests the RNA strand, the complementary cDNA strand synthesis is initiated at the polypurine tract (PPT) just inside the 3'LTR. With the double stranded DNA copy complete, the *pol*-derived integrase protein cuts the genomic DNA at the target site with nicks offset by 3-5 base pairs (bp). The insertion of the new retrotransposon copy and filling of the gaps caused by the nicks produces a new TE insertion with characteristic 3-5bp flanking sequence duplications at the insertion site (BOEKE AND CORCES 1989).

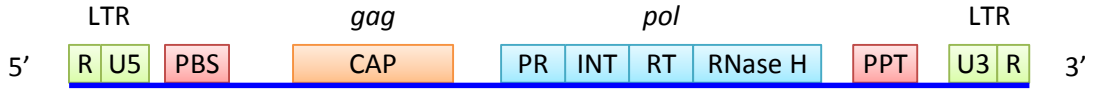


Figure 2.3 Structural arrangement of the mRNA intermediate of a Ty1-*Copia* retrotransposon. Key to elements is the same as Figure 2.2. Diagram is not drawn to scale.

Dating retrotransposons

Transposition requires TE-encoded proteins that are correctly transcribed, translated, and processed into functional enzymes. Over time, random mutations that accumulate within TEs destroy their capacity for autonomous transposition and they join the bulk of inactive or non-autonomous repetitive elements in the genome. Screening TE copies for those with complete open-reading frames (ORFs) is therefore one way to identify the most recently active autonomous copies.

Since the mobilisation of LTR-retrotransposons involves duplication of the direct repeats at both ends, paired LTRs are identical immediately following insertion. As DNA copy errors and mutations accumulate over time, LTR sequences gradually diverge. The sequence identity between an LTR pair therefore begins at 100% and decreases over time. As a result, LTR sequence identity can be considered to be molecular clock, indicative of the age of a given insertion (VITTE *et al.* 2007).

2.2.3 Insertion site preferences

The chromatin material within the nucleus of a living cell is a complex and dynamic molecular arrangement of genomic DNA, protein and RNA. Chromatin can be morphologically divided into two types. Regions of chromosomes that are in a relaxed state, capable of binding with proteins associated with RNA transcription, are termed euchromatin. Heterochromatin, conversely, is tightly condensed in the interphase nucleus and can be distinguished by darker histochemical staining (for review see (VAN STEENSEL 2011).

Genic sequences that are not being actively transcribed are maintained in a condensed state. These regions, which differ between tissues within an organism, are described as functional heterochromatin. In contrast, constitutive heterochromatin is necessary for the structure, function and integrity of chromosomes. These patterns of condensed chromatin are stable across generation boundaries. They include centromeres, which attach to spindle fibres during karyokinesis, and telomeres, which protect the integrity of genic regions close to DNA ends.

Chromosome-scale analyses of TE dispersion patterns show that these elements are largely localised to constitutive heterochromatic regions, where they often appear as tandemly repeated, and even nested, insertions (SANMIGUEL *et al.* 1996). The presence of few TEs in the euchromatic regions of

most genomes led some to propose an insertion affinity towards heterochromatin (SANMIGUEL *et al.* 1996). However, novel insertions appear, in most cases, to target actively transcribed genes (CRESSE *et al.* 1995; LIU *et al.* 2009; NAITO *et al.* 2009; RAKOCEVIC *et al.* 2009; ITO *et al.* 2011). The discrepancy between distribution patterns for new and ancient insertions is likely due to greater negative selection pressures on elements that alter gene expression patterns (LISCH AND BENNETZEN 2011). Additional support for this hypothesis comes from the study of the tobacco retrotransposon *Tnt1*. New insertions appear preferentially within or close to coding sequences, but are not removed with time. This is proposed to be due to the polyploid nature of the tobacco genome, which results in a relaxation of negative selection pressures (LE *et al.* 2007).

A couple of active TEs in plants (Tam1, a class II element from Antirrhinum and Ttd1a, a class I element from durum wheat) have nevertheless been identified which do appear to insert mainly into regions of repetitive DNA (NACKEN *et al.* 1991; WOODROW *et al.* 2010).

2.2.4 Epigenetic regulation of transposition

Active transposons can rapidly increase in number and greatly change the genomic landscape (NAITO *et al.* 2009; SCHNABLE *et al.* 2009). Organisms therefore require strict silencing techniques to prevent the deleterious effects of TE mobility on the genome.

Inspired by her studies of the Ac/Ds loci in maize, Barbara McClintock proposed that modifications to the nuclear DNA that do not directly alter the nucleotide sequence could affect the expression of genes (MCCLINTOCK 1951). Such changes, which have collectively come to be called epigenetic modifications, are now known to include chromatin remodelling by nucleic acid-binding proteins and methylation of cytosine bases. Though all somatic cells of an organism contain a complete copy of the genome, epigenetic modification patterns limit the spatial and temporal production of proteins, resulting in different and specialised cell types.

The ability for epigenetic modifications to be altered also provides the genome with a degree of responsiveness to changing conditions. The seasonal timing of flowering to spring, which has been best studied in *Arabidopsis*, is one example of this. In *Arabidopsis*, flowering is ordinarily repressed by FLC, a MADS-box DNA binding protein. When the plant endures a protracted period of cold (vernalisation), an increasing number of cells accumulate epigenetic modifications that block the expression of FLC. This 'memory' of cold exposure is maintained through multiple successive rounds of mitotic division, allowing the flowering process to complete. Inhibition of FLC expression is not maintained through meiosis, however, so that that the following generation will once again wait for cold conditions before flowering (KIM *et al.* 2009; ANGEL *et al.* 2011).

Several systems that are involved in this suppression of gene expression also act to prevent transposon activity. The following sections give a description of these.

Post-transcriptional silencing

The eukaryotic RNA interference (RNAi) system actively removes the mRNA of repetitive sequences from the transcriptome in a sequence-specific manner. As LTR-retrotransposons increase in copy number nested insertions become more common (SANMIGUEL *et al.* 1996). The generation of transcripts by RNA polymerase II from these conformations can produce sense-antisense complementary strands, which anneal to form double stranded RNA (dsRNA). The DICER family of proteins cleave dsRNA into 21-30nt short interfering RNA (siRNA) fragments. These fragments guide an RNA-induced silencing complex (RISC) to target transcripts. Identified target molecules are then degraded by ARGONAUTE (AGO) proteins within the complex (FESCHOTTE *et al.* 2002; BAULCOMBE 2004; SLOTKIN AND MARTIENSSEN 2007).

Chromatin modifications

The methylation states of DNA-bound histone proteins affect chromatin folding, and thereby the availability for sequences to be transcribed. The vast majority of TEs in the genome are maintained in a non-transcriptionally active state within heterochromatin. Three signals of transcriptionally inactive chromatin, dimethylation of histone H3 at lysine 9 (H3K9me2), methylation at lysine 27 (H3K27me1) and methylation of histone H4 at lysine 20 (H4K20me1), have been found to be enriched in TEs, but not genic regions (GENDREL *et al.* 2002; ROUDIER *et al.* 2011).

The conversion of relaxed euchromatin to a condensed state by histone modification is also triggered by sequence-specific siRNAs. An RNA induced transcriptional gene-silencing (RITS) protein complex is targeted to nascent transcripts still attached to RNA Pol II. Digestion of the nascent transcripts by Argonaute proteins in the RITS complex is believed to recruit H3K9 methyltransferase to the transcription site. The resulting histone modifications of nearby DNA cause condensation of the chromatin (LIPPMAN AND MARTIENSSEN 2004; VERDEL *et al.* 2004; GREWAL 2010).

The deacetylation of H3 and H4 histones by HISTONE DEACETYLASE 6 (HDA6), through interaction with METHYLTRANSFERASE 1 (MET1), has also been associated with the suppression of TE activity (LIU *et al.* 2011). Likewise, *Arabidopsis* plants lacking a functional copy of the chromatin-remodelling enzyme DECREASE IN DNA METHYLATION 1 (DDM1) have been found to lose H3K9 chromatin modifications. Both MET1 and DDM1 mutations lead to reactivation of endogenous TEs (MIURA *et al.* 2001; GENDREL *et al.* 2002; LIU *et al.* 2011). The precise actions that DNA methyltransferases play in chromatin modification are still unclear, however.

DNA methylation

A second type of epigenetic mark, the methylation of cytosine bases, renders DNA unavailable as a template for transcription by RNA polymerase II. Unlike in animals, DNA methylation in plants is almost exclusive to transposons and other repetitive DNA (LIPPMAN *et al.* 2004).

To maintain stable suppression of transposon activity, methylation patterns must be copied with each replication of the DNA. Symmetrical methylation patterns, which include CG and CHG (where H = C, A or T), can be copied between hemimethylated strands. Methylation at CG dinucleotides is maintained by MET1, but the process requires the presence of DDM1. Likewise, methylation of CHG sites is maintained by the plant-specific CHROMOMETHYLASE 3 (CMT3), but depends on a H3K9 methyltransferase partner, KRYPTONITE (KYP) (BARTEE *et al.* 2001; CHAN *et al.* 2005). It is clear therefore that the processes of epigenetic silencing by methylation of DNA and histones are related.

Asymmetrical methylation patterns (i.e. CHH) are present only on one DNA strand, and as such are lost on one of the new DNA copies following replication. These marks are constantly re-established through the process of RNA-dependant DNA methylation (RdDM). This system starts with the plant-specific RNA polymerase IV (RNA pol IV), which is able to transcribe heterochromatic regions of the DNA (HERR *et al.* 2005). Second-strand synthesis of RNA pol IV transcripts is performed by RNA-dependent RNA polymerase (RDR) proteins, (XIE *et al.* 2004). It is still not clear how target RNA sequences are selected for processing, but transcripts with multiple microRNA sites appear to act as a trigger (AXTELL *et al.* 2006). Dicer-like3 (DCL3) digests the dsRNA molecules into 24-nt siRNAs, which are recruited by AGO4 and AGO6 in a pathway that parallels RNAi. The siRNA-loaded AGO4/6 proteins recognise complementary nascent transcripts of DNA-dependant RNA polymerase V (RNA pol V). The methyltransferase DRM2 is then recruited to methylate cytosine bases at the RNA pol V transcription site (LAW AND JACOBSEN 2010; NUTHIKATTU *et al.* 2013). It therefore appears that transposable elements that are not correctly transcribed for expression and mobilisation are nevertheless capable of stimulating siRNA production and are still involved in the maintenance of silencing.

The Pol IV-initiated RdDM pathway is dependant on existing heterochromatic copies of repetitive DNA to maintain methylation of these elements. This leaves the paradox of how TEs initially come to be methylated. A second RdDM pathway has been identified which is not involved in the maintenance of TE silencing, but instead appears to act as a link between the posttranscriptional gene silencing / RNAi pathway and DNA methylation.

In this pathway, an RNA-dependant RNA polymerase and a DICER-like enzyme, named RDR6 and DCL2 respectively, produce 21-22nt siRNAs from the Pol II mRNA transcripts of actively transcribed TEs. These siRNAs trigger the same AGO6 and Pol V proteins described above, which act as effectors

of DNA methylation. This alternative initiation pathway is thought to be responsible for the silencing of new and recently reactivated TEs, thereby controlling the generation of new heterochromatin (LAW AND JACOBSEN 2010; NUTHIKATTU *et al.* 2013) .

Trans silencing

Not all silenced plant transposons are found in nested arrangements or as tandemly repeated insertions. The ability of the RNA-dependant silencing mechanisms to scan the genome for sequence identity means that a single active TE can trigger *trans* silencing of all related sequences (GIRARD AND HANNON 2008). Eukaryotic genomes are able therefore to recognise, silence and inactivate TEs with remarkable efficiency and precision. The same systems are also employed for transgene silencing (CHAN *et al.* 2005). The aptitude of host genomes towards mobile DNA suppression has led to the hypothesis that gene-silencing mechanics evolved as a necessary response to the activity of transposons and viruses, and were later adopted for the regulation of endogenous genes (FLAVELL 1994; SLOTKIN AND MARTIENSSEN 2007; LISCH AND BENNETZEN 2011).

2.2.5 Transposition as a response to a changing environment

The systems described above are essential for limiting the genotoxic effects of mobile DNA. Yet, while the vast majority of transposition events are detrimental (as is the case for most mutations in general), TE insertions can occasionally provide adaptive advantages. Being mostly sessile, plants are particularly exposed to local environmental changes. When faced with an existential threat, the slender chance of a favourable mutation greatly increases in value, while the need to eschew mutation of vital alleles remains constant.

Abiotic stresses

A few years after McClintock's early TE work, Harrison and Fincham (1964) described the instability of the PAL locus of *Antirrhinum majus,* which causes streaked colour patterns in the petals. This phenotype has since been found to be caused by a class II transposon, *Tam3*, which is active only at environmental temperatures below 15°C (CARPENTER *et al.* 1987). The temperature-sensitvity is due to a nuclear localization inhibitory domain in the *Tam3* transposase, which interacts with host factors to prevent nuclear import at higher temperatures (FUJINO *et al.* 2011). In an interesting contrast to *Tam3*, a TE in rice named *ONSEN* only shows activity at high temperatures (ITO *et al.* 2011).

Experiments by Virginia Walbot showed that the "genomic stress" of maize tissue exposed to gamma irradiation was linked to the activation of *Mutator* TEs (WALBOT 1988). The same TEs are even active following irradiation with UV light calibrated to simulate field conditions with 33% ozone depletion (WALBOT 1992a; WALBOT 1999). Recently, a retrotransposon in Durum wheat, *Tdt1a*, has been found

to be stimulated even by high levels of light in the visible spectrum, as well as salt stress (WOODROW *et al.* 2010).

Biotic stresses

Pouteau and co-workers (1994) found that the activity of the tobacco retrotransposon *Tnt1* is generally very low in plants, but increases markedly during protoplast isolation. The use of a *Trichoderma* crude extract for cell wall digestion was determined to trigger the activity of these elements. Moreover, the necrosis-inducing extracts of other fungal pathogens were also found to increase transcription of *Tnt1*. Further experiments showed that *Tnt1* activity increases in response to viral infection and a variety of abiotic stresses, including cold shock and wounding (MHIRI *et al.* 1997). The transcription of a second tobacco retrotransposon, *Tto1*, is similarly induced by fungal elicitors, plant wounding and tissue culture (TAKEDA *et al.* 1999).

Tissue culture

Epigenetic changes to DNA of plant tissues in culture appear to allow the reactivation of multiple transposon classes. In rice, the frequencies of new insertions of one retrotransposon family (Tos) and one class II TE family (nDiaZ) have been found to increase with tissue culture period (HIROCHIKA *et al.* 1996; HUANG *et al.* 2009). The *MERE1* retrotransposon family is also active during tissue culture in at least two accessions of *Medicago truncatula*. The transcriptional activity of *MERE1* is correlated with a decrease in cytosine methylation across the element, and leads to a high (40%) rate of novel insertions in regenerated plants, most of which are within genes (RAKOCEVIC *et al.* 2009).

Cis-regulatory elements

Cis-regulatory elements (CREs) associated with defence response pathways have been identified in several of the plant retrotransposons activated by stress conditions. Short repeated elements were identified in the LTRs of the tobacco transposons *Tnt1* and *Tto1* that are similar to sequences in defence-response gene promoters. Testing of these motifs with transgenic reporter gene constructs revealed that they act as enhancers, increasing expression in response to wounding, tissue culture and other abiotic and biotic elicitors (GRANDBASTIEN 1998; TAKEDA *et al.* 1999).

In addition to defence-response CREs, the light-activated *Ttd1a* element of durum wheat contains the TCCC-motif associated with light-responsive expression of the spinach plastocyanin gene (WOODROW *et al.* 2010). Other *cis*-elements associated with defence, tissue-specificity and light responsiveness have been found in sunflower retrotransposons (VUKICH *et al.* 2009). Temperature-regulated expression of TEs has so far been associated with epigenetic change (TITTEL-ELMER *et al.* 2010), or interaction of TE proteins with host factors (FUJINO *et al.* 2011), rather than CREs.

2.2.6 It's not all bad news: How TEs pay the rent

Throughout the history of their study, transposons have been defined as parasitic DNA (ORGEL AND CRICK 1980; SABOT AND SCHULMAN 2006; GIRARD AND HANNON 2008). Yet increasingly, examples have been found which may justify the ubiquitous preservation of TEs within genotypes.

Maintenance of chromosome structure

Transposons constitute the majority of pericentromeric and telomeric heterochromatin in most genomes (SLOTKIN AND MARTIENSSEN 2007). Sequence data show that the centromeric retrotransposon (CR) lineage of Ty3-*Gypsy* retrotransposons, and their typical association with centromeric satellite repeats, pre-date the divergence of the monocots and eudicots (DU *et al.* 2010). Aside from acting as functional constituents of these chromosomal structures, it appears that TEs throughout the genome may be associated with maintaining the constitutive heterochromatin in a condensed state. In *Arabidopsis*, 24-nucleotide siRNAs produced by TEs have been associated with centromere methylation and those produced by centromeric satellite repeats have likewise been linked to TE silencing (MAY *et al.* 2005).

The non-coding telomeres, which lie at the end of eukaryotic chromosomes, are thought to protect the euchromatic regions from erosion. They are composed of tandemly repetitive DNA, added via the reverse transcription of RNA templates by the enzyme telomerase. The similarity of telomerase in both structure and mechanism to the reverse transcriptase genes of non-LTR retrotransposons is thought to be due to their descent from a common ancestral sequence (LINGNER *et al.* 1997). In fact, in *Drosophila*, which does not have a telomerase homologue, the function of telomerase is replaced by two non-LTR retrotransposons *HeT-A* and *TART*, which accumulate at the chromosome ends in a developmentally controlled manner (PARDUE AND DEBARYSHE 2011).

Gene creation

By transporting fragments of endogenous genes during their mobilisation, TEs rearrange the genome, shuffling exon fragments and creating new ORFs. New sequence arrangements are subject to selective pressures, those that persist contribute to genome evolution (KAZAZIAN 2004; FESCHOTTE AND PRITHAM 2007).

An analysis of Mutator-like elements in rice revealed the ability of some of these elements to capture and transport portions of host genes throughout the genome as they transpose. Approximately one-fifth of over 3000 of these so-called Pack-MULEs were found to be carrying exons from multiple genome locations. Regions of over a thousand genes appear to have been mobilised in this way (JIANG *et al.* 2004). Pack-MULEs have since also been found in the genomes of maize and *Arabidopsis*,

and a comparison of these elements across the three genomes reveals the preferential capture of CG-rich gene fragments (FERGUSON AND JIANG 2011).

Helitrons, unlike most class II TEs, are not flanked by TIRs and are therefore difficult to detect in sequence data. Their specific sequence type is associated with an unusual form of rolling-circle transposition, however. Over successive rounds of 'cut-and-paste' mobilisation, these elements collect small regions from ORFs at their insertion sites. Transcripts of both pack-MULES and Helitrons have been observed that consist of captured exons correctly spliced together (BENNETZEN 2005). It is therefore highly likely that these TE types contribute directly to new gene creation.

Similar genomic restructuring has occurred in grapevine through the activity of MITEs (BENJAK *et al.* 2009). One element family, *mPifvine-3.1,* is particularly common in the 3' untranslated region (UTR) of genes. The authors that identified this family propose that this location bias is due to an unknown selective advantage conferred by the elements, rather than insertion site specificity.

Gene regulation

Sequence motifs within retrotransposon LTRs regulate mRNA synthesis from the transcription start sits (TSS) within the 5'LTR (see Figure 2.2). Being identical in sequence, the 3'LTR can drive the formation of mRNA transcripts from sequence downstream of the TE insertion, potentially placing genes under the regulatory control of the TE. Sequence similarity between the CREs of an active tobacco *Tto1* retrotransposon and a plant defence gene (AoPR1) indicate that an ancient *Tto1* insertion may possibly have contributed some stress-response regulatory motifs to the promoter of this gene (TAKEDA *et al.* 1999). As much as one quarter of all promoter regions in the human genome contain TE-derived sequences, including regulatory motifs (JORDAN *et al.* 2003).

Ito and colleagues (2011) showed that heat shock activation of the *Arabidopsis* retrotransposon *ONSEN* results in new insertions in plants deficient of siRNA biogenesis. Genes neighbouring new *ONSEN* insertions also became heat responsive. A study by Naito and colleagues (2009) has even shown the preferential insertion of the stress-activated rice DNA transposon mPing into gene promoters, resulting in a large proportion of genes at insertion loci coming under stress-responsive transcriptional regulation. Through selection, therefore, TE-derived sequences become domesticated as functional and necessary parts of the eukaryotic genomes.

TE 'immunity'

As described (see 2.2.4), TEs that are transcriptionally dormant as a result of epigenetic modifications are still capable of contributing siRNAs. In this way, TEs maintain asymmetrical DNA methylation patterns and provide a trigger for silencing new insertions. The maintenance of a diverse library of TE types within the genome therefore facilitates the rapid identification and silencing of new potentially

mutagenic elements. In addition to a collection of reference elements, it is possible that RDR proteins are involved in the maintenance of a pool of dsRNA, enabling an even swifter response to the reactivation of dormant transposons.

Active TE demethylation

Plants lacking the systems required to maintain epigenetic modifications lose DNA methylation at each replication cycle, in a process known as passive demethylation (TEIXEIRA AND COLOT 2010). A subfamily of DNA glycosylases have been identified in *Arabidopsis*, however, which actively remove methylated cytosine bases from the DNA, leaving a gap which is filled with unmethylated cytosine (ZHU 2009).

One of these enzymes in particular, DEMETER (DME), is specifically expressed during plant female gametogenesis in the central cell, which gives rise to the triploid endosperm following fertilisation (CHOI *et al.* 2002). An increase in DME expression leads to genome-wide demethylation and a non-CG hypermethylation at siRNA-targeted loci. In contrast, the DNA of the embryo itself is hypermethylated, with a particularly high CHH methylation across transposable elements. The authors propose that transposon methylation may be removed in the non-gametic endosperm in order to allow the activation of transposons and a consequent accumulation of TE-targeting siRNAs. Trafficking of these siRNAs to the embryo may be responsible for enhanced silencing of mobile elements (HSIEH *et al.* 2009).

Slotkin and colleagues (2009) drew the same conclusions about the mechanism for TE silencing in gametes after studying male gametogenesis in *Arabidopsis*. Their experiments showed an apparent downregulation of the DDM1 gene in vegetative nucleus of pollen, accompanied by reactivation of certain TEs. The same TEs are hypermethylated in adjacent sperm cells, where siRNAs corresponding to the specific TEs abound.

2.2.7 Using TEs as molecular markers

Most of the initial interest in TE polymorphism has been in regards to their use as molecular markers for genotyping. Large (usually 5-10kb) TE insertions produce dominant markers that are easy to detect. Class I transposons, in particular, do not catalyse their own excision, and so the reversion of an insertion locus is unlikely. The dynamic nature of these elements makes them particularly suitable as genetic markers, since their mutation rates are often above the background rates of SNP accumulation and microsatellite replication slippage. Several PCR-based techniques have been proposed to date for the identification and screening of the insertions of polymorphic TE families. The specifics of these, and their comparative efficacies are described in more detail later (See 4.2.2).

2.2.8 TEs as tools for mutagenesis

TEs are a natural source of mutations within organisms, capable of rapidly changing the genomic landscape. For example, the sequence data of the two maize genomes sequenced to date differ by approximately 22%, of which 90% is attributable to TE-derived sequences (VIELLE-CALZADA *et al.* 2009).

The mutagenic capacity of endogenous TEs has been used to generate mutation saturated populations in multiple plant species, including *Antirrhinum majus* (CARPENTER AND COEN 1990), rice (HIROCHIKA 1997), maize (WALBOT 2000) and lotus (FUKAI *et al.* 2012; URBAŃSKI *et al.* 2012). As discussed (see 2.2.4), the silencing apparatus of host genomes is well adapted to suppress transposon activity. In order to overcome the restriction that this places on mutagenesis experiments, researchers have introduced foreign TEs by transgenesis. Saturation mutagenised populations have successfully be produced using the maize *Ac/Ds* DNA transposon system in rice (KOLESNIK *et al.* 2004) and barley (KOLESNIK *et al.* 2004), the maize *En/Spm* system in *Arabidopsis* (TISSIER *et al.* 1999) and the tobacco retrotransposon tobacco *Tnt1* in *M. truncatula* (TADEGE *et al.* 2008).

2.2.9 TEs in grapevine

The 504.6Mb estimated size of the grapevine genome (VELASCO *et al.* 2007) is relatively small when compared with other angiosperms. Nevertheless, TEs are estimated to compose between 21.5% and 41.4% of the total sequence data (JAILLON *et al.* 2007; VELASCO *et al.* 2007). Using next-generation sequencing technologies, genomic variations have been compared between clones. Of the mutation types contributing to inter-clonal genetic disparity (SNPs, indels and TE insertion polymorphisms), TE insertions represent the largest proportion (CARRIER *et al.* 2012).

Class I transposable elements

Prior to the public release of the assembled grapevine genome, three grapevine retrotransposon families had been identified. Two of these, *Vine1* (VERRIES *et al.* 2000) and *Gret1* (KOBAYASHI *et al.* 2004) were serendipitously uncovered by researchers attempting to characterise mutant alleles of unrelated genes. A third, *Tvv1* was revealed by a chromosome walking experiment which began with degenerate primers targeting conserved regions within the reverse transcriptase domains of Ty1-*Copia* retrotransposons (PELSY AND MERDINOGLU 2002).

In 1999 a British research group described a PCR-based approach which utilises biotinylated probes to capture retrotransposon sequence fragments from purified DNA (PEARCE *et al.* 1999). The probes are designed to specifically target the M1 conserved motif in the RNase H domain of retrotransposons. Following removal of unlabelled fragments, a second round of PCR utilises nested primers that target a second motif, M2. Enriched amplification products are then cloned and

sequenced. Moisy and colleagues (2008) applied this technique to Pinot noir genomic DNA, and obtained partial sequence data for ten new Ty1-*Copia* retrotransposons. By aligning the recovered fragments to the grapevine genome data, they were able to extract full-length sequences for each of the new retrotransposon families and characterise the distribution of these families in the reference genome. Copies of elements containing complete ORFs were found for three of the families, indicating that these elements may still be autonomously active.

Class II transposable elements

Since the transposition of class II transposons begins with their excision from the host genome, mobilisation of these elements does not generally result in an increase in copy number. While illegitimate recombination events can raise copy numbers over time, this class of element does not have the same tendency as retrotransposons to rapidly proliferate. Consequently, just 1.98% of the *V. vinifera* genome is composed of class II TEs (BENJAK *et al.* 2008). An *in silico* search of the two published grapevine genomes uncovered a total of 3246 DNA transposons belonging to four out of five of the class II TE superfamilies known to exist in plants: *CACTA*, *hAT*, *Mutator* and *PIF* (BENJAK *et al.* 2008). Interestingly, as many as 1160 of these elements appear to be potentially intact, half of which are from the *hAT* superfamily. Examples of each superfamily have been found with intact transposase ORFs, which may indicate potential for transposition. Furthermore, examples of each of the four superfamilies identified have been found in expressed-sequence tag (EST) databases (BENJAK *et al.* 2008).

The following year, the same group published an analysis of MITEs (see 2.2.1) in the grapevine genome. Their conclusions were that MITEs have been highly mobile during the history of grape domestication and breeding, contributing significantly towards the genetic diversity among varieties (BENJAK *et al.* 2009).

A single example of the contemporary activation of a TE in grapevine has been described to date. The reiterated reproductive meristem (RRM) somatic variant of the Carignan grapevine variety, which displays extreme flower cluster proliferation, is a result of a class II TE, named *hatvine1-rrm*. An insertion of this element in the promoter of the VvTFL1A gene causes a dominant mutation that results in the overexpression of the gene, although not in a constitutive manner. The TE ORF is inserted in reverse orientation with respect to VvTFL1A, and expression of the gene is still developmentally regulated, suggesting that the mutation may have resulted in the disruption of a repressor site (FERNANDEZ *et al.* 2010). This example illustrates the positive and complex effects that TE insertions can have on the expression of adjacent genes.

2.2.10 The *Gret1* insertion in VvMYBA1

The berries of all wild grapes turn red upon ripening, as an attractant to animals, which eat the berries and inadvertently act as agents of seed dispersal (SLINKARD AND SINGLETON 1984). This change occurs during a period of ripening known as *veraison*, when water-soluble anthocyanin pigments are produced in berry skins, where they accumulate, changing the colour of the berries.

The production of anthocyanin is inhibited in white grapes by mutations in two adjacent, paralogous transcription factor genes. The insertion of a retrotransposon named *Gret1* (for *G*rape *ret*rotransposon *1*) in the promoter of VvMYBA1 prevents assembly of the transcription activation complex at this site, inhibiting transcription (KOBAYASHI *et al.* 2004). In addition, a non-conservative SNP in the gene VvMybA2 causes an altered and truncated amino acid sequence (WALKER *et al.* 2007). Without either of these transcription factors the final enzyme of the anthocyanin biosynthetic pathway, UDP-glucose: flavonoid 3-O-glucosyltransferase (UFGT), is not produced. Grapes lacking UFGT ripen, but remain green in colour.

Though the white grape phenotype would probably be rapidly eliminated by natural selection, the same mutations are found in all white table and wine grape varieties, indicating that human agriculture has preserved the trait by active selection and propagation (LIJAVETZKY *et al.* 2006; FOURNIER-LEVEL *et al.* 2010). This is an explicit example of the potential of TEs to produce mutant phenotypes that are of value to the wine industry, and to agriculture in general.

2.3 Grapevine Somatic Embryogenesis

Due to the limitations of heterozygosity and generation times (as discussed in 2.1.2), propagation and regeneration of new plantlets *in vitro* from tissue culture material has become central to the improvement of woody perennial crop plants. However, such species are often particularly recalcitrant to adventitious bud formation. Mullins and Srinivasan (1976) were the first to report the successful culture of totipotent grapevine tissue, and the regeneration of whole plantlets from this tissue. These researchers found that ovules of the Cabernet Sauvignon variety, cultivated on media containing added auxin (β-naphtoxyacetic acid; NOA) and cytokinin (6-benzyl amino purine; BAP) could produce stable embryogenic callus (EC), from which plantlets could be germinated by altering phytohormone concentrations in the media.

In recent years, interest in the *in vitro* regeneration of grapevine plantlets has increased as transgenic techniques have acquired a record of successful crop improvement. Many groups have attempted to regenerate plantlets from grapevine tissue. However, efforts have been frustrated by the variable degrees of success obtained with different grapevine varieties.

2.3.1 Issues of chimerism

Initial grapevine transformation experiments showed that introduction of foreign DNA via Agrobacterium is possible. However, the use of shoot apex cultures as starting material produced plant shoots that contained both transformed and non-transformed cells (BARIBAULT *et al.* 1990). The chimerism of recovered plants is likely to be a result of regeneration from partially transformed material that is able to survive low levels of antibiotic selection. A stepwise increase in antibiotic concentration in regenerating plant media has been successful in limiting chimerism somewhat with respect to transgenes following transformation by organogenesis (MEZZETTI *et al.* 2002). However, the chimeric nature of meristem remains a hurdle to the regeneration of phenotypically homogenous vines from this tissue (see section 2.1.4)

2.3.2 Somatic embryo tissue culture

Regeneration of grapevines from embryo cultures does not preserve the periclinal chimerism of the parent plant (FRANKS *et al.* 2002; BERTSCH *et al.* 2005). As a result, somatic embryos have been the tissue type of choice for transformation experiments. Since the initial report by Mullins and Srinivasan, attempts have been made to initiate somatic embryo cultures from a host of grapevine varieties, as described in the following sections. While many researchers have tested a wide array of media types, explant types and growth conditions, still no single protocol has been achieved which is universally applicable to *V. vinifera*. Hormone concentrations, the growth stage selected for explant

material collection and even agar types have been reported as decisive factors in achieving the consistently low rates of successful somatic embryogenesis.

Explant material

Floral tissues have been almost exclusively used for initiation of somatic embryo cultures. Of these, the excised anthers of unopened flowers are most common (KIKKERT *et al.* 1997; TORREGROSA 1998; IOCCO *et al.* 2001).

While attempting to adapt the initiation of embryogenic callus (EC) from anthers to a variety of different cultivars, Perrin and colleagues (2004) found anthers cut from the calyx rather than plucked were twice as efficient for the initiation of EC, as initiation always begins at the cut site. The stage of flower development also contributes significantly to successful initiation, with most successful initiations being from stage II (6-8cm flower clusters with 1.5mm diameter buds and 0.8-1.0mm long, yellowish, translucent anthers) or stage III (9-10cm flower clusters with 1.5-2.0mm diameter buds and 1.0mm long, yellowish, cloudy anthers) flowers (DHEKNEY *et al.* 2009).

Carimi and co-workers (2005) reported higher EC initiation efficiencies from stigma and style tissues than from anthers for three grapevine cultivars, although it should be noted that only one stigma-style unit is obtained per dissected flower, as opposed to five anthers. Kikkert and colleagues (2005) similarly found a 7-fold average increase in initiation efficiency when using ovaries in comparison with anthers for the initiation of EC from twelve *Vitis* genotypes. Whole flowers have also been found to be appropriate for EC initiation in four cultivars (GAMBINO *et al.* 2007).

A couple of research groups have reported the generation of EC from the young leaves of certain grapevine varieties, but this technique has not yet proved successful for any of the famous wine-making varieties (DAS *et al.* 2002; DHEKNEY *et al.* 2009).

Culture regimes

The success and efficiency of initiation of EC is dependant on both the nutrient and phytohormone concentrations of media used. Several cultivars have been successfully initiated on PIV medium, which uses major elements according to Nitsch and Nitsch (1969) with 4.5µM 2,4-Dichlorophenoxyacetic acid (2,4-D) and 8.9µM BAP (FRANKS *et al.* 1998; IOCCO *et al.* 2001). Torregrosa (1998) reported successful initiation of EC from four grapevine varieties using C_1^P medium, which contains major elements according to Murashige and Skoog (1962) at half-strength, 5µM 2,4-D and 1µM BAP. Both media types use microelements according to Murashige and Skoog. Using a three-step series of media, with different phytohormone concentrations at each step, Perrin and colleagues (2004) successfully initiated EC from anthers of 19 grapevine genotypes on medium containing macro-and microelements according to Murashige and Skoog. Fifteen of these could be

maintained as established EC for at least 84 days. More recently, the generation of somatic embryos has been reported for many more grapevine varieties, but a single broadly applicable technique has still not been identified (DHEKNEY *et al.* 2009; OLAH *et al.* 2009).

While EC cultures are commonly maintained with periodic subculturing on the same solid media used for initiation, there is evidence that moving EC onto media with higher ammonium concentrations favours proliferation (PERRIN *et al.* 2001). Long-term maintenance of EC as suspension cultures in liquid medium has also been reported (MAURO *et al.* 1995). Ben Amar (2007) and co-workers showed that establishment of liquid cultures from EC masses is improved by the use of medium pre-conditioned with previously established EC cultures or through the addition of arabinogalactan.

2.4 Open questions

The literature reviewed here shows that in the past half-century much momentum has been built in elucidating the true nature and function of the non-coding regions of the genome. The pace has been accelerated by recent technological advances that have improved the speed and economy of DNA sequencing and epigenetic DNA characterisation. In grapevine, somatic embryogenesis techniques have enabled the rapid regeneration of non-chimeric individuals from vegetative material, providing an optimal tissue in which to study somatic mutagenesis.

At this juncture several interesting and important questions are apparent: Which of the manifold endogenous TEs that have proliferated within the grapevine ancestral genome remain active? To what extent is transposon mobility responsible for somatic mutation in vines? Do aspects of *terroir*, such as environmental events and endemic microflora, affect rates of TE mutagenesis? Could the capability to manipulate and track rates of transposition in grapevine be used to produce new grapevine clones?

The experiments presented in the following chapters have been designed to address these questions.

Chapter 3

An *in silico* Analysis of Grapevine LTR-Retrotransposons

3.1 Overview

The publication of entire draft genome sequences to publicly available databases has been taken as an invitation by many researchers to move from traditional gene-by-gene genetics to the broader field of genomics. In the past, the characterisation of mobile elements has generally been triggered by the discovery of specific mutant alleles. But by using ever-increasing computational resources to analyse whole genome data, it is now possible to study the entire mobilome – the full complement of mobile DNA in the genome.

In this chapter *in silico* techniques are used to gain a broad perspective of the retrotransposon portion of the grapevine mobilome. The abundance and integrity of retrotransposon insertions are investigated in the context of distinct, repetitive families. Retrotransposon sequence data are used to compare the phylogeny of these elements and their classification is discussed with regards to previously defined clades. The internal regions of five retrotransposon families (*Gret1*, *Edel*, *Cremant*, *Noble* and *Tvv1*) were selected for further analysis. The long terminal repeats (LTRs) of these elements were screened for regulatory sequence motifs to determine conditions under which they may be transcriptionally active. The abundance of retrotransposon families in publicly available grapevine expressed sequence tag databases was also determined, as an indication of which families may retain transcription potential.

These data show that each of the retrotransposon clades known to exist in plants is represented in grapevine. Certain retrotransposon families have historically undergone great proliferation and endure as thousands of eroded fragments. A few families are composed of largely complete elements, and may retain transpositional functionality under certain conditions. The LTR sequences of these elements indicate a stress-responsive pattern of activity.

3.2 Introduction

3.2.1 Retrotransposon-derived sequences in the grapevine genome

The large size of most plant genomes is due in the most part to the proliferation of class I transposable elements (TEs). Though the 505Mb genome of domesticated grapevine is relatively small among angiosperms, an estimated 19.4% - 23.5% of the total genome is made up of retrotransposon-derived sequences, of which 96.7% - 96.8% are attributable to LTR-retrotransposons. In comparison, only an estimated 1.4% - 1.7% of the genome is composed of class II transposons (VELASCO *et al.* 2007).

Using a software tool called ReAS (LI *et al.* 2005), Jaillon and colleagues (JAILLON *et al.* 2007) assembled high-depth next generation sequencing (NGS) reads rejected by genome assembly software to reconstruct consensus sequences representative of ancestral repetitive DNA elements. The assembled elements have been deposited in Repbase update, an online database of repetitive elements identified in eukaryotic genomes (JURKA *et al.* 2005). In addition, ten grapevine retrotransposon families were defined following their amplification from grapevine DNA with degenerate PCR primers (MOISY *et al.* 2008).

To characterise and compare the retrotransposon families present in grapevine, the canonical sequences in Repbase were used as query sequences to retrieve individual insertions from the PN40024 grapevine genome sequence data. This was performed using two different software algorithms. The BLAST algorithm was used to retrieve sequences varying in similarity and coverage, with respect to the Repbase consensus sequences, from the grapevine reference genome. The BLAT algorithm (KENT 2002) performs a more rapid search than BLAST, but retrieves only sequences with very high similarity (above 95%). This tool was used to retrieve recent insertions (i.e. almost complete elements with high sequence similarity).

Sequence alignment and phylogeny software were used to analyse the number of insertions, integrity of elements and both the intra- and interfamilial sequence divergences of the repetitive elements recovered in order to compare the historical proliferation and degradation of the retrotransposon families.

3.2.2 LTRs as promoter elements

The expression of eukaryotic genes is regulated by the assembly of transcription factor proteins and RNA polymerase II at conserved sites in the promoter area upstream of an ORF. In this region *cis*-regulatory element (CRE) motifs serve as binding targets for specific transcription factors. Once bound to the DNA, transcription factors either promote or hinder transcription. In the case of

retrotransposons, the 5' LTR acts as the promoter, housing the transcription start site (TSS) and CREs. The LTR sequences also act as functional elements however, such that their direct sequence identity is ensured at each mobilisation of event. As a result, the 3' LTR of a retrotransposon can initiate transcription of sequences downstream of the element (Figure 3.1). Depending on the location of the retrotransposon in the genomic context, this may lead to the production of aberrant mRNAs or the co-regulation of endogenous genes by the retrotransposon.

Figure 3.1 Transcription is initiated within the LTR sequences of retrotransposons.

After comparing retrotransposon families, previously characterised CRE motifs were identified in the LTR sequences of five retrotransposons, using the plant cis-acting regulatory DNA elements database (PLACE; (HIGO *et al.* 1999), in order to obtain information about the conditions under which these elements may be transcriptionally active. To obtain empirical information about the transcriptional activity of these elements, the NCBI's collection of grapevine expressed sequence tag (EST) databases was searched for retrotransposon sequences. The CRE variation, EST transcript abundance, and insertion numbers in the genome suggest highly variable patterns of activity across retrotransposon families.

3.3 Methods

3.3.1 Extracting retrotransposon insertions from the PN40024 genome

Canonical retrotransposons were reconstructed from LTR and internal element sequences stored in the Repbase Update database (Version 18.05; http://www.girinst.org/repbase/). All *V. vinifera* repeat sequences were downloaded and simple sequence repeats (SSRs) in the data were masked with N's using RepeatMasker (version 4.0.2; http://www.repeatmasker.org). The names assigned by Moisy and colleagues (2008) were used in place of Repbase nomenclature where applicable. Full elements were reconstructed by reassembling the sequences for each family in the format: LTR-Internal-LTR.

The reconstructed retrotransposons were used to search a local copy of the 12X PN40024 grapevine genome using the blastn algorithm of the BLAST+ software package (version 2.2.28; available from http://blast.ncbi.nlm.nih.gov/). For each element, hits with less than 80% identity or less than 20% coverage of the query sequence were discarded. Those that remained were grouped into six categories according to their coverage of the query TE sequence: 20-50%, 50-70%, 70-80%, 80-90%, 90-100% and 100%+). The *V. vinifera* chloroplast genome (from the Maxxa variety) and the mitochondrial genome (from the ENTAV115 clone of Pinot noir) were also searched in this way, as well as with MegaBLAST (ZHANG *et al.* 2000) and RepeatMasker.

To determine the number of complete retrotransposon copies for each family present in the genome, the genome sequence was queried with the reconstructed canonical elements using the BLAT search program (KENT 2002) via the URGI server (http://urgi.versailles.inra.fr/mobyle/cgi-bin/portal.py). Of the returned hits, those within 10% of the query sequence size were counted to ensure ancient TE fragments were not included.

3.3.2 Grapevine retrotransposon phylogeny

Long terminal repeat sequences in the grapevine genome were identified by querying a local BLAST database of the assembled PN40024 genome data with the Repbase LTR elements, using the MegaBLAST algorithm. Hits with greater than 80% identity and 50% coverage of the query sequence were aligned using the MUSCLE computer program (EDGAR 2004). Neighbour-joining distance trees were drawn using the HKY genetic distance model (HASEGAWA *et al.* 1985) and branch stability was tested by 100 bootstrap iterations with 70% support threshold.

Sequence data of 548 Ty1-*Copia* and Ty3-*Gypsy* retrotransposons characterised during assembly of nine different angiosperm genomes (apple, *A. thaliana*, grape, *M. truncatula,* papaya, peach, pear, soya bean and woodland strawberry) were downloaded from Repbase. The internal sequences of the

elements were aligned separately using ClustalW (THOMPSON *et al.* 1994). Sequence similarities were compared using 100X bootstrapped neighbour-joining consensus trees.

3.3.3 Regulatory motif screening in LTR sequences

Based on the copy number results and reports of previously active retrotransposons in grapevine (see 2.2.9), five retrotransposon families were selected for further sequence analysis. Transcription start sites (TSS) prediction was performed on these elements using the TSSP-TCM software (SHAHMURADOV *et al.* 2005) available online at http://linux1.softberry.com/berry.phtml?topic=tssp&group=programs&subgroup=promoter). Sequences corresponding to the 5'LTR plus 5' untranslated region (UTR) of these retrotransposon families were searched for known *cis*-regulatory elements using the New PLACE online search tool (https://sogo.dna.affrc.go.jp/cgi-bin/sogo.cgi?sid=&lang=ja&pj=640&action=page&page=newplace). Transcription start sites, TATA-box motifs and CREs that have been previously associated with stress response were annotated on the sequences using Geneious software (version 6; www.geneious.com). The BLAST+ program suite was used to search the NCBI *V. vinifera* EST databases for transcripts matching the internal portions of each grape retrotransposon family.

3.4 Results

3.4.1 The distribution and integrity of retrotransposon insertions in the grapevine genome

Defined retrotransposon families

The Repbase Update database contains complete sequences of 102 Ty1-*Copia* and 35 Ty3-*Gypsy* elements from *V. vinifera* (Table 3.1). For each element, the LTR and internal sequences have separate entries, generally named in the format 'Gypsy1-VV_LTR' and 'Gypsy1-VV_I' respectively. Four of the 137 families (*Gret1*, *Tvv1*, *V1* and *GYVIT1*) were deposited individually by the researchers who identified and sequenced these elements. The remainder were deposited by the international research collaboration that published the PN40024 grapevine genome sequence (JAILLON *et al.* 2007).

Table 3.1 Numbers of *V. vinifera* repeat sequences in Repbase by superfamily.

	Complete	Partial / non-autonomous
Class I		
Ty1-*Copia*	102	0
Ty3-*Gypsy*	35	1
LINE	10	0
Caulimovirus-like	3	0
Class II		
En/Spm (CACTA)	7	3
Harbinger (PIF)	4	7
hAT	9	3
MuDR	14	4
Helitron	1	0
Unclassified	2	
Total	187	18

Retrotransposon-derived sequences in the grapevine genome

Searches of the grapevine reference genome identified a total of 14,460 sequence fragments derived from the Ty1-*Copia* and Ty3-*Gypsy* retrotransposon superfamilies (Table 3.2). Large proportions of the insertions from each superfamily (48.4% for Ty1-*Copia* and 77.3% for Ty3-*Gypsy*) were less than one-fifth of the size of complete elements of their respective families. Six Ty1-*Copia* and eight Ty3-*Gypsy* families are very highly represented with respect to the other retrotransposon families, accounting for 43.2% and 65.1% of the total insertions for their respective superfamilies. However, only 1.1% of all elements in these families are at least 90% complete (Figure 3.2 and Figure 3.3). Two Ty1-*Copia* families, *Noble* and *Edel*, are most notable for their high numbers of complete elements (Figure 3.2). The four retrotransposon families which are studied in further detail in later chapters (*Gret*, *Edel*, *Cremant* and *Nobel*) are indicated specifically on Figure 3.2 to Figure 3.5.

Table 3.2 Integrity of Ty1-Copia and Ty3-Gypsy fragments in the PN40024 grapevine reference genome. Calculated as a per cent coverage of the respective Repbase consensus sequences.

TE Integrity	Ty1-Copia	Ty3-Gypsy
20%+ complete	3,424	5,215
50%+ complete	1,281	1,663
70%+ complete	990	248
80%+ complete	620	180
90%+ complete	284	16
100%+ complete	493	46
Total	7,092	7,368

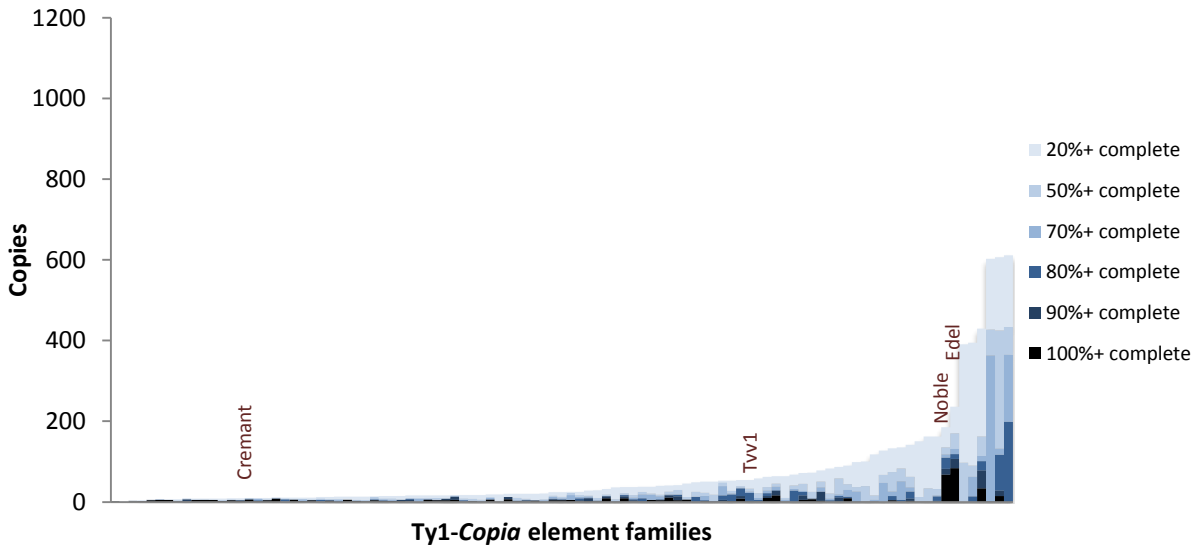


Figure 3.2 Ty1-*Copia* retrotransposon-derived sequences in the grapevine genome. All 102 Ty1-*Copia* families are shown along the x-axis. Shading indicates sequence integrity with respect to full elements. Families studied in further detail in the following sections are named.

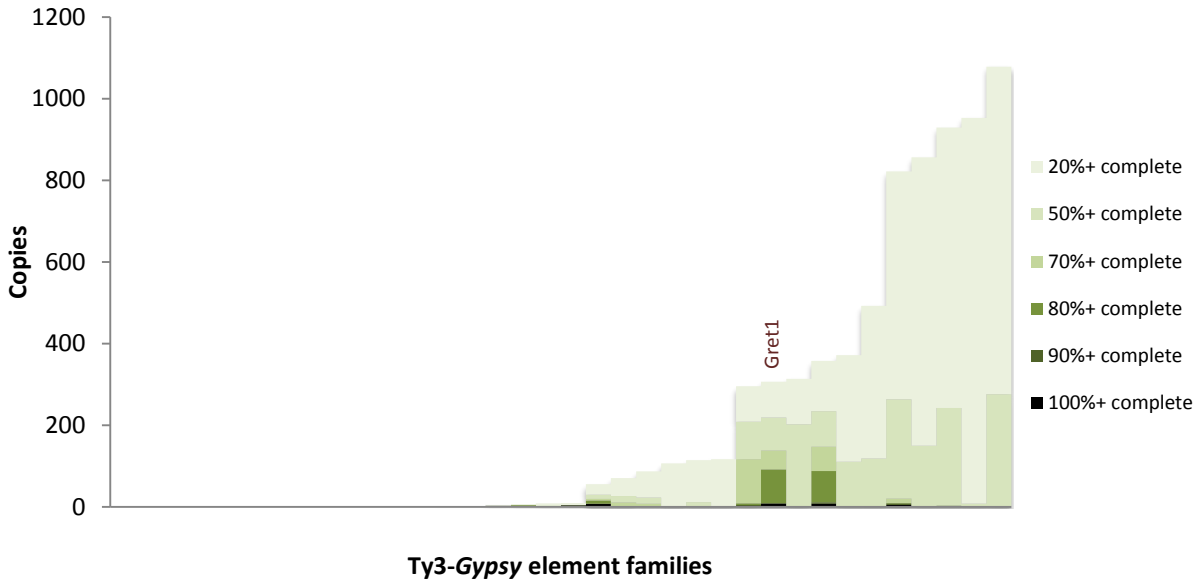


Figure 3.3 Ty3-*Gypsy* retrotransposon-derived sequences in the grapevine genome. All 35 Ty3-*Gypsy* families are shown along the x-axis. Shading indicates sequence integrity with respect to full elements. The *Gret1* family (studied in further detail in the following sections) is indicated.

No complete TEs were found in the plastid genomes. Three small (less than 60bp) fragments with sequence similarity (greater than 70% identity) to class II TEs were found in the chloroplast (two EnSpm-5_VV and one EnSpm-4N1_VV). Several (102) short sequence fragments (104-450bp) in the mitochondrial genome have greater than 65% identity to internal regions of retrotransposon sequences, but none with similarity to any of the Repbase LTRs were identified. No retrotransposon family is represented more than once in these mitochondrial genome fragments.

Complete retrotransposon insertions

The BLAT search program uses a rather different algorithm to BLAST to locate sequences with a much higher identity threshold (greater than 90%) at a faster rate. Within the Ty1-*Copia* superfamily, the number of full insertions (within 10% of the size of the canonical Repbase sequences) returned by BLAT search on the PN40024 genome averaged 11.9 per family. Three families had particularly large numbers of full elements (Figure 3.4). These same three families had the highest numbers of mostly complete elements (90%+) in the BLAST search data (Figure 3.2). The Ty3-*Gypsy* superfamily showed a more evenly distributed range of insertion numbers, with an average of 10.6 insertions per family (Figure 3.5).

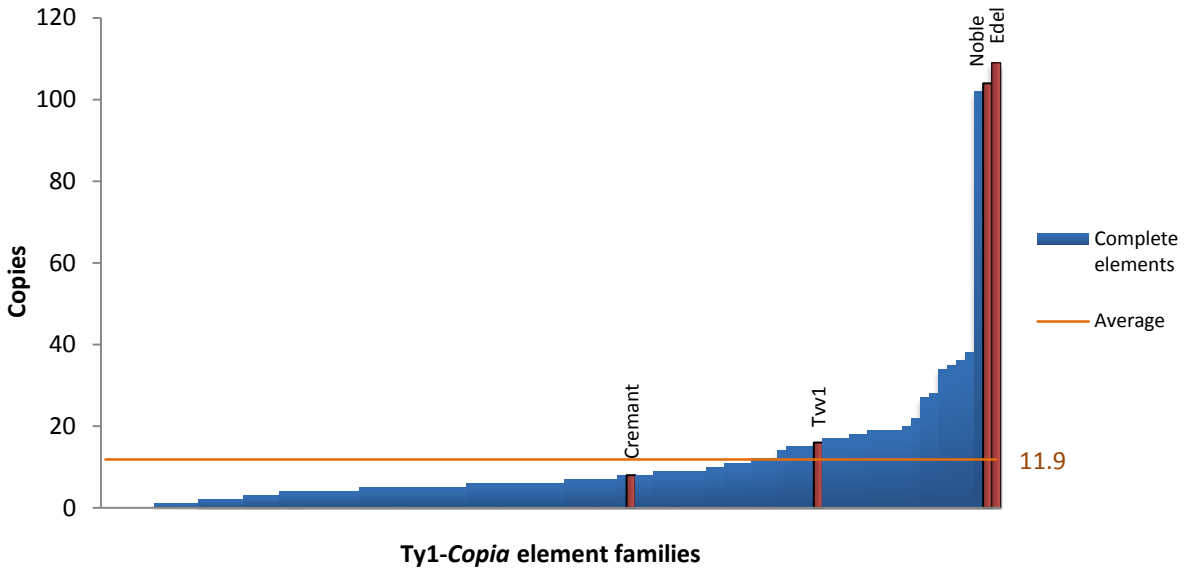


Figure 3.4 Number of complete copies for Ty1-*Copia* retrotransposon families in the grapevine genome. All BLAT search hits in the 12X PN40024 genome assembly with > 90% identity and > 90% coverage were counted. Transposons families studied in later chapters appear in red.

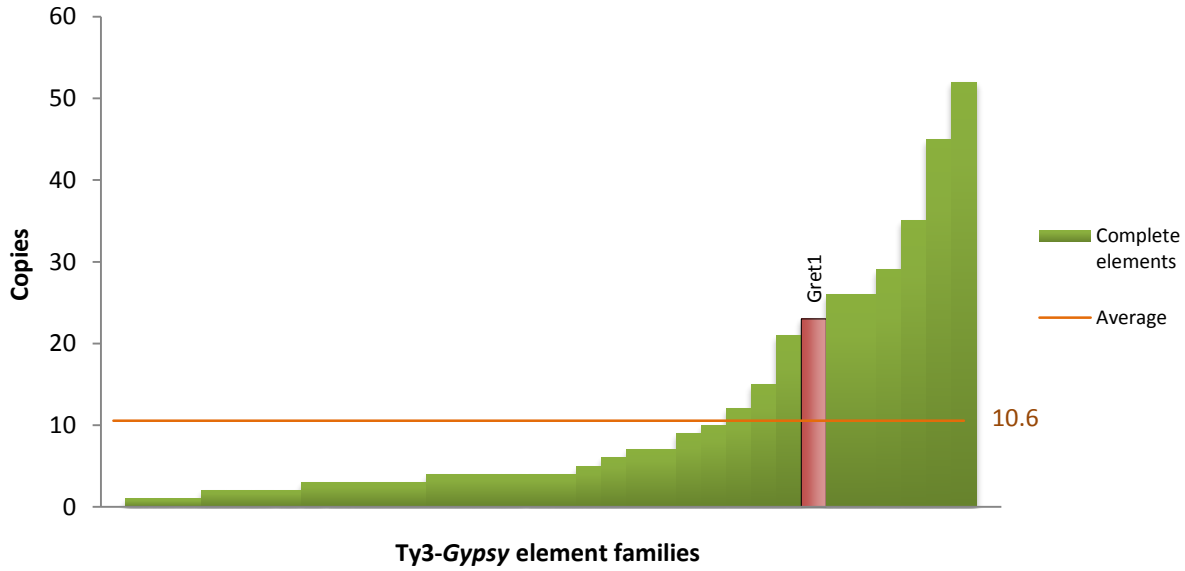


Figure 3.5 Number of complete copies for Ty3-*Gypsy* retrotransposon families in the grapevine genome. All BLAT search hits in the 12X PN40024 genome assembly with > 90% identity and > 90% coverage were counted. Transposons families studied in later chapters appear in red.

The completeness of all retrotransposon-derived sequence fragments (i.e. the size ratio of every TE fragment with respect to the full size element from which it is derived) reveals the historical degradation of ancient insertions within the genome. Figure 3.6, which represents all retrotransposon-derived fragments with greater than 90% identity to the canonical Repbase sequences, shows an exponential increase in the number of insertions as fragment size decreases.

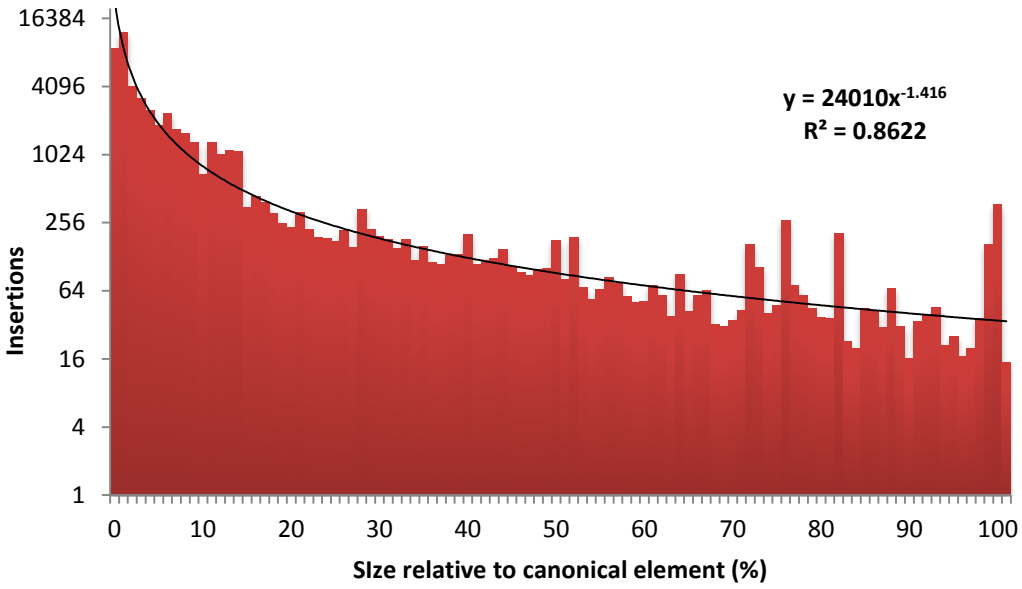


Figure 3.6 Integrity of retrotransposon-derived sequences in the grapevine genome. Total number of insertions for all combined retrotransposon families in the 12X PN40024 *V. vinifera* genome, grouped by size percentile relative to full-length elements. Y-axis scale is logarithmic.

3.4.2 Phylogeny of grape retrotransposons

Sequence similarity between families

The sequence variation between the LTRs of canonical retrotransposons is high, particularly within the Ty1-*Copia* superfamily. As a result of the lack of sequence commonality between families, only a few clades could be defined (Figure 3.7).

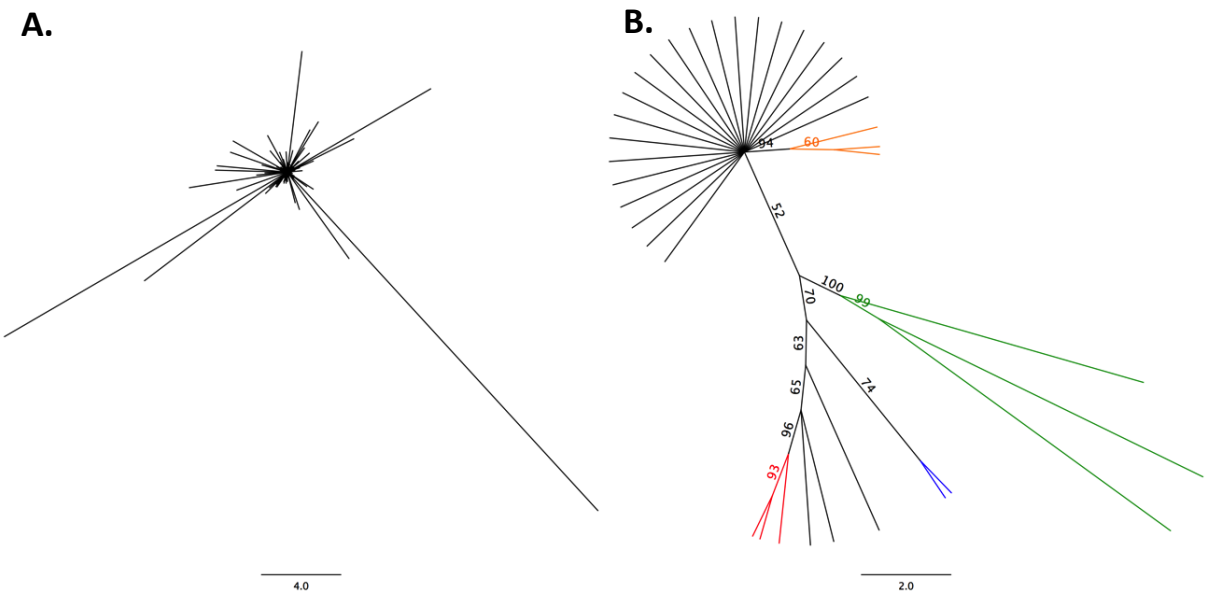


Figure 3.7 Phylogenetic comparison of grapevine LTR-retrotransposon families using LTR sequence data. (**A**) Cladogram of the Ty1-*Copia* superfamily; (**B**) Cladogram of the Ty3-*Gypsy* superfamily.

Clades common to Figure 3.8 are shown in the same colour on that image. Bootstrap values (100X) are indicated.

The retrotransposon internal sequences (i.e. the sequence portion between LTR flanks) proved much more useful for comparing relatedness between families (Figure 3.8). The few retrotransposon families that could be grouped according to their LTR sequences also associated most closely when their respective internal sequences are compared (highlighted in the same colours on Figure 3.7 and Figure 3.8).

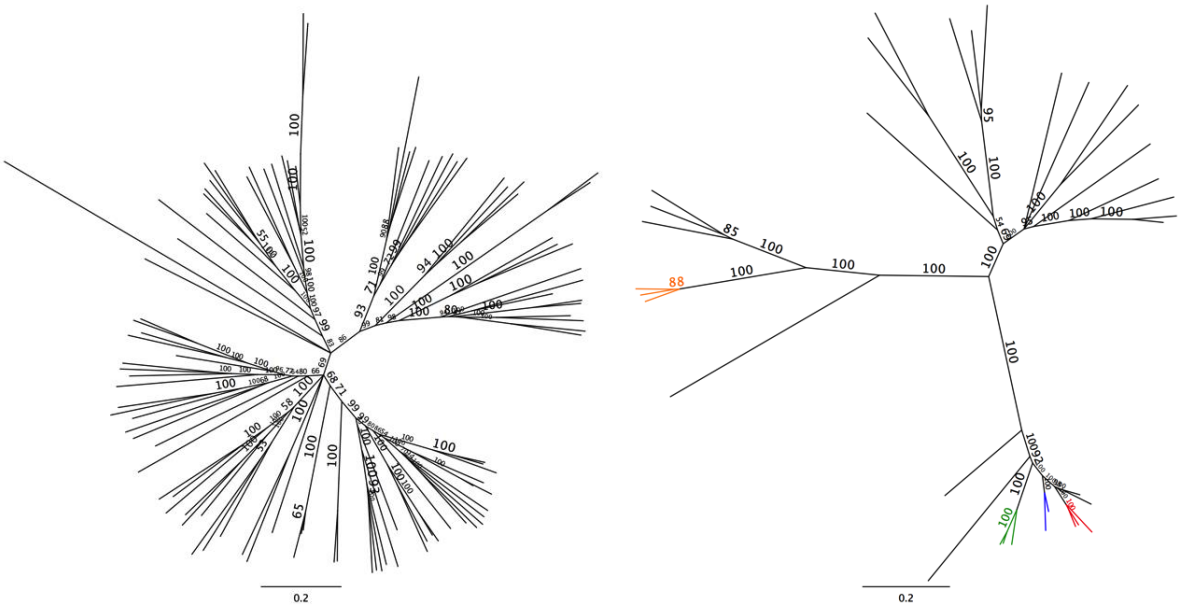


Figure 3.8 Phylogenetic comparison of grapevine LTR-retrotransposon families using internal sequence data. (**A**) Cladogram of the Ty1-*Copia* superfamily; (**B**) Cladogram of the Ty3-*Gypsy* superfamily. Clades common to Figure 3.7 are shown in the same colour on that image. Bootstrap values (100X) are indicated.

When elements from other species that have previously been used to define retrotransposon clades are included in the sequence alignments, it is possible to match branches of the grapevine retrotransposon trees to those clades that have been previously described in other organisms (Figure 3.9).

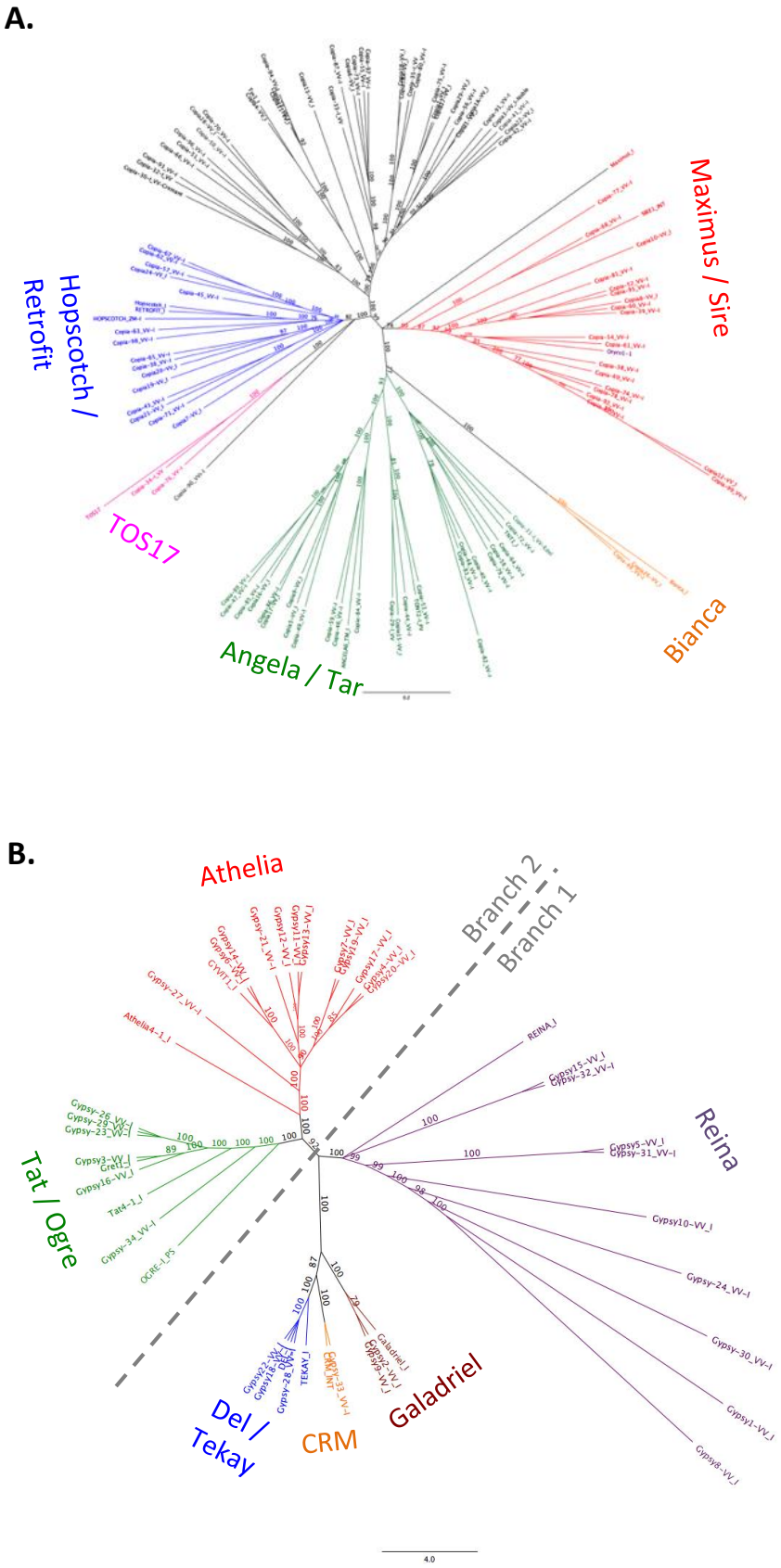


Figure 3.9 Organisation of grapevine retrotransposons according to previously reported clades (see Table 2.1). (A) Ty1-*Copia* superfamily (B) Ty3-*Gypsy* superfamily. Alignments were produced using ClustalW and trees were drawn using the Geneious Tree Builder according to the HKY genetic distance model.

Alignments of internal sequences between the flanking LTRs of retrotransposons were used to compare the relatedness of retrotransposons from nine angiosperm species (Figure 3.10). Although retrotransposon families of particular species (most notably *Arabidopsis thaliana* and *Glycine max*) do form a number of species-specific clades, this is not the rule for most of the genomes analysed, including *V. vinifera* (Figure 3.11 and Figure 3.12). In these genomes, retrotransposon families frequently show higher similarity to elements from other species than to their nearest intraspecific relatives.

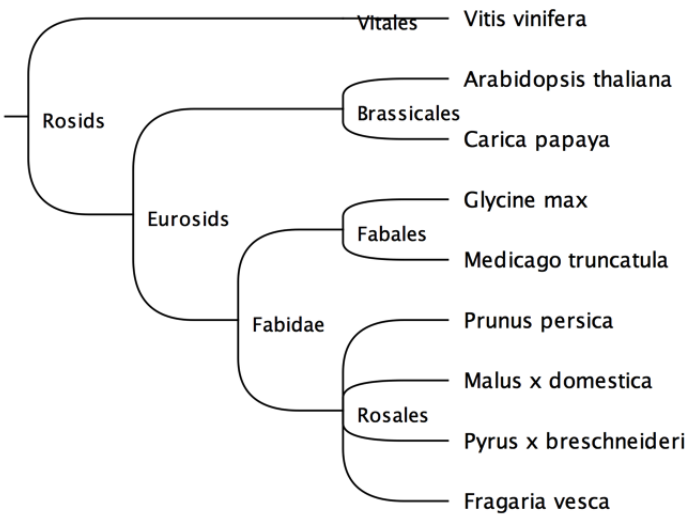


Figure 3.10 Phylogenetic relationships among the nine plant species used for retrotransposon sequence comparisons. Phylogeny is according to Wang and co-workers (2009).

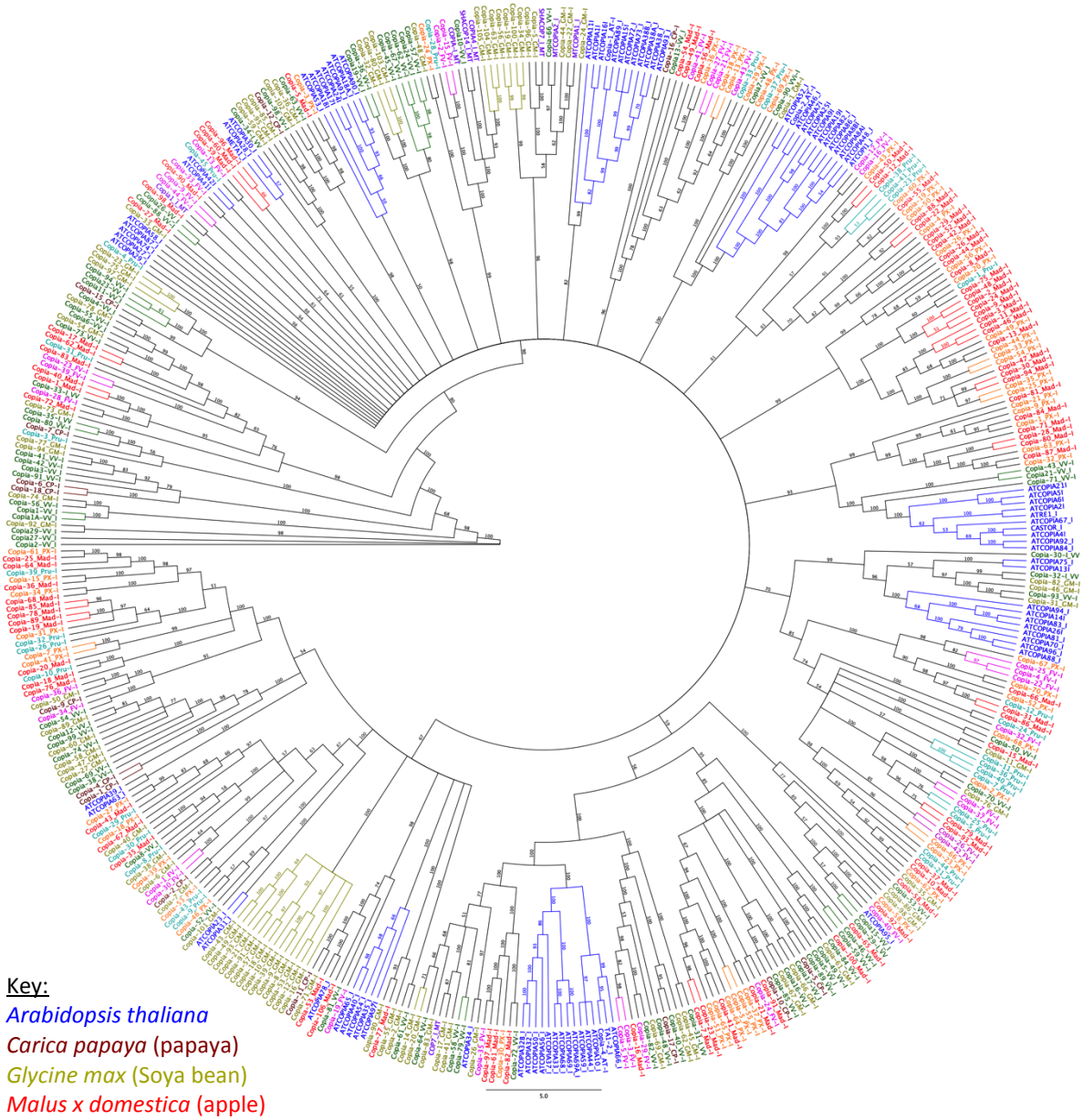


Figure 3.11 Sequence identity of 469 Ty1-*Copia* retrotransposons from nine angiosperm species. The consensus tree is based on the alignment of regions internal to the flanking LTRs, constructed from 100 bootstrap replicates with a 70% support threshold. TE families are colour coded by species.

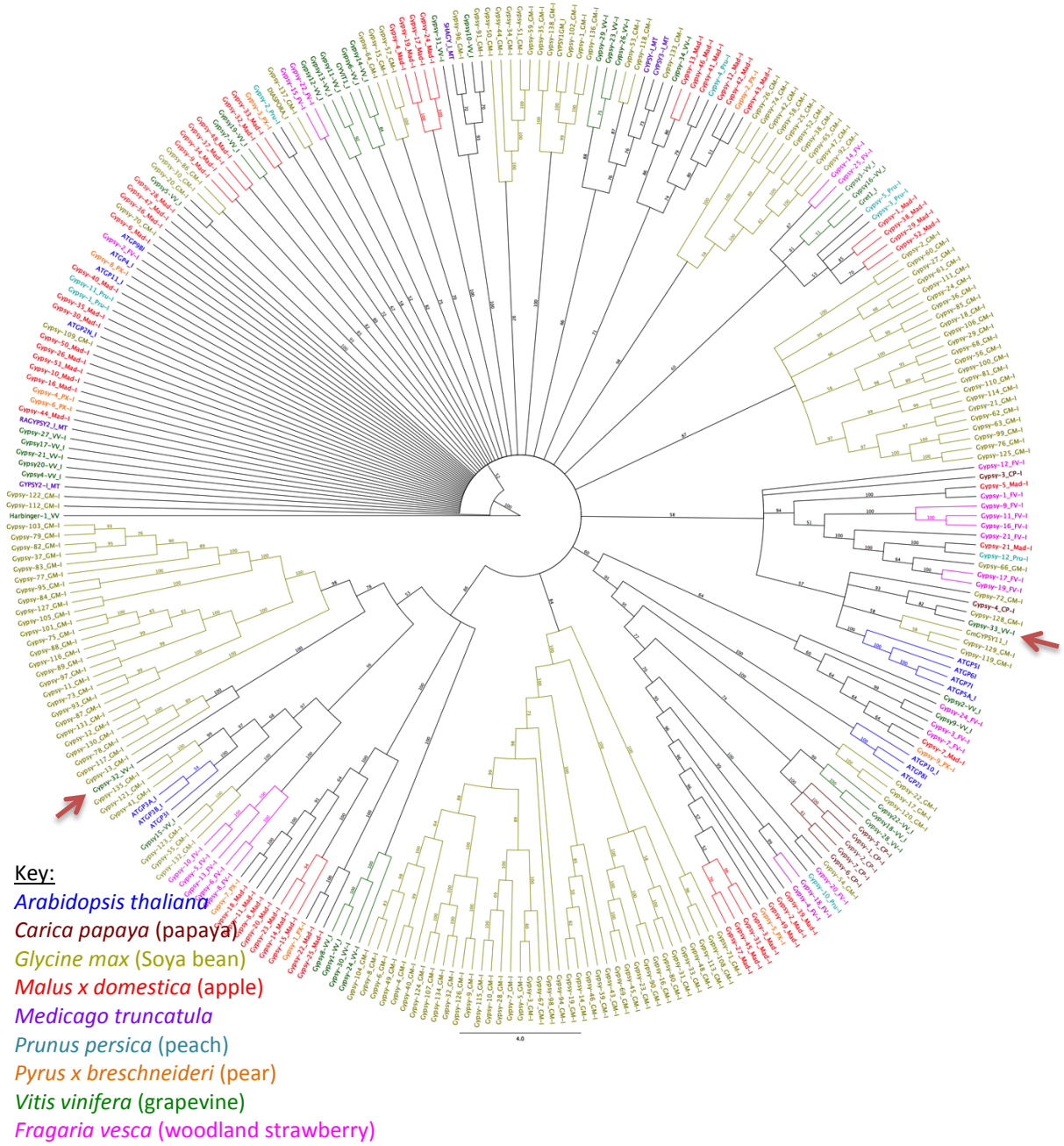


Figure 3.12 Sequence identity of 296 Ty3-*Gypsy* retrotransposons from nine angiosperm species. The consensus tree is based on the alignment of regions internal to the flanking LTRs, constructed from 100 bootstrap replicates with a 70% support threshold. TE families are colour coded by species. Examples of grapevine retrotransposon families with high sequence similarity to elements from other species are indicated by red arrows.

Sequence similarity within families

In contrast with the multiple sequence alignments of canonical elements representing distinct families, the LTR sequences of individual retrotransposon insertions within families showed high sequence identity (Appendix B.1). Since only a small number of loci differentiate sequence pairs, few clades remained after bootstrap analysis was applied to trees produced from these alignments (Figure 3.13). Note that the divergence scale (substitutions per site) is two to three orders of

magnitude lower than that observed between families (Figure 3.7). Nevertheless, several stable clades were observed among the LTRs of certain families, of which *Gret1* is an example.

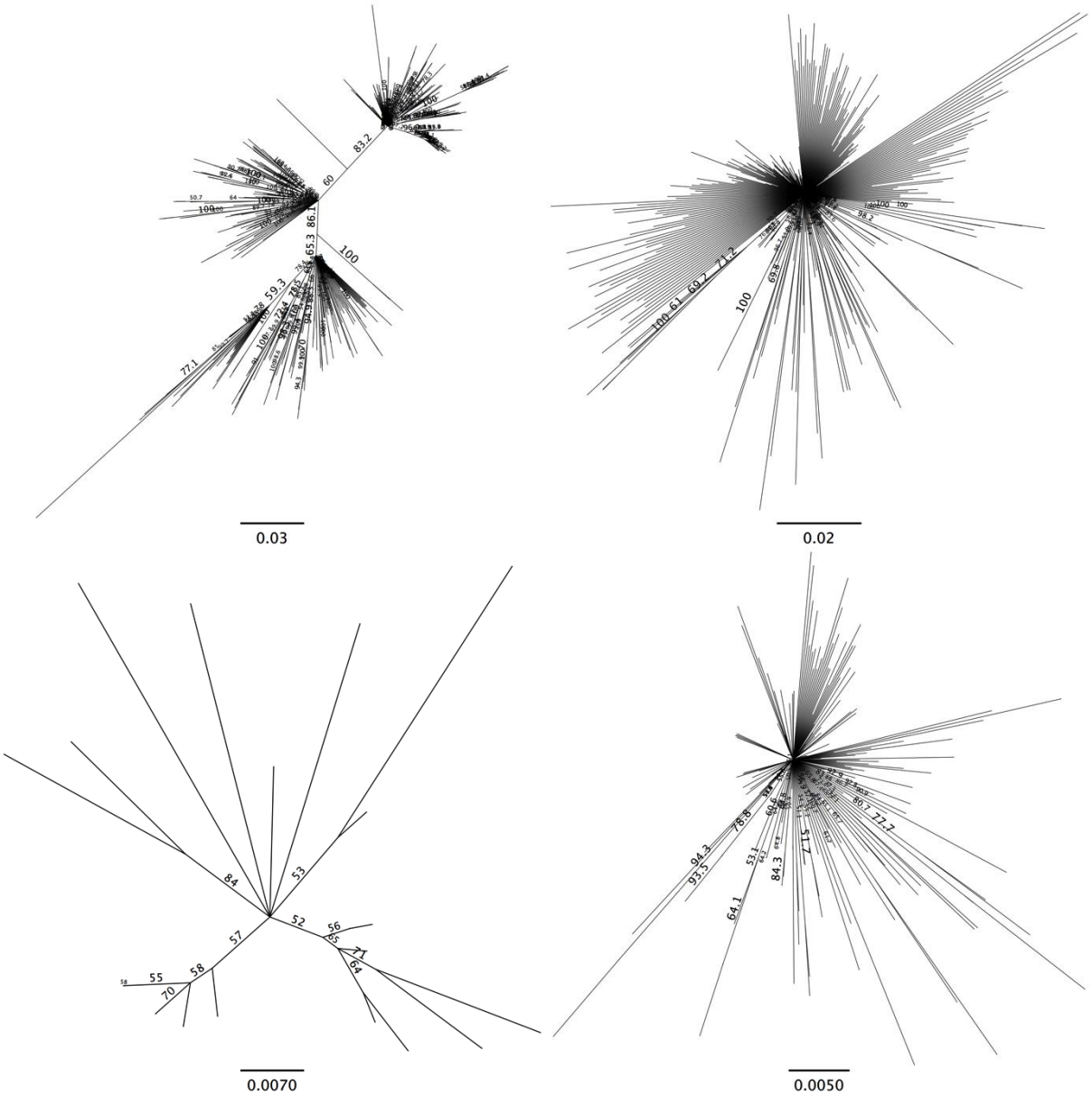


Figure 3.13 Neighbour-Joining trees of LTR insertions for four retrotransposon families
(Constructed from 100X bootstrap replicates with 50% support threshold; scale bar shows substitutions per site; branch stability is labelled on the diagrams). Clockwise from top-left: *Gret1*, *Edel*, *Noble, Cremant*.

Long terminal repeat regions were used for these comparisons because their higher relative sequence divergence makes them more informative than internal regions. Additionally, multiple sequence alignments of internal sequences require a great deal of computational resource due to their large sizes (approximately 5-10kb) and the large numbers of insertions within certain families

(upwards of 5,000 elements) makes pairwise comparisons of these regions prohibitive. The bootstrapped cladograms produced from intrafamilial LTR alignments do not provide enough information to describe the lineages of most insertions. However, these diagrams do present the overall structure of each family, including the number of elements and major clades within each family, in a simple, visual format. The LTR cladograms of all grapevine retrotransposon families are given in Appendix B.2, Table B.1.

3.4.3 Stress-response motifs in five retrotransposon families

Sequence motifs known to be associated with stress response were found in each of the retrotransposon sequences analysed (Table 3.3). The majority of those CREs identified have been associated with dehydration response. Other notable associated stresses include temperature shock, pathogen response and wound response. The 15bp HBOXCONSENSUSPVCHS motif found 152bp upstream of the TSS of *Gret1* is also present in *Tnt1*, a tobacco retrotransposon known to become active in response to wounding and pathogen attack. Regulatory elements were also identified in *Gret1* and *Edel* that have been associated with seed or meristematic tissue-specific expression.

Table 3.3 Cis-regulatory elements associated with stress or tissue-specific expression present in LTRs and 5'UTRs of five retrotransposon families (K=G/T; N=A/C/G/T; R=A/G; W=A/T; Y=C/T; ABA: abscisic acid). Motifs that have been linked experimentally with pathogen response in plant systems are highlighted in red. The average number of occurrences for each element in the promoters of eight characterised *V. vinifera* pathogen response genes (PR10.2 - PR10.9) is included in the final column for comparison. Expression association shown is that listed in each curated PLACE record.

Cis-regulatory element (PLACE database name)	Sequence	Expression association	Counts per retrotransposon sequence					PR gene Average
			Gret1	*Edel*	*Cremant*	*Noble*	*Tvv1*	
ABREATCONSENSUS	YACGTGGC	general stress / ABA response	1	0	0	0	0	0.0
ABREATRD22	RYACGTGGYR	dehydration response	1	0	0	0	0	0.0
ABRELATERD1	ACGTG	dehydration/ etiolation response	5	0	1	2	2	0.3
ACGTATERD1	ACGT	dehydration/ etiolation response	9	2	1	2	6	0.5
AGCBOXNPGLB	AGCCGCC	pathogen response	1	0	0	0	0	0.4
ASF1MOTIFCAMV	TGACG	general stress/ xenobiotic/ auxin response	2	1	0	1	2	0.4
CACGTGMOTIF	CACGTG	G-box; defence response; embryogenesis-specific expression	2	0	0	2	0	0.1
CBFHV	RYCGAC	dehydration response	0	0	0	3	2	0.3
CCAATBOX1	CAAT	heat-shock	0	0	0	2	0	3.1
CRTDREHVCBF2	GTCGAC	temperature/ dehydration response	0	0	0	1	0	0.0
DRECRTCOREAT	RCCGAC	temperature/ salinity/ dehydration response	0	0	0	2	1	0.3
EMBP1TAEM	CACGTGGC	general stress / ABA response	1	0	0	0	0	0.0
GCCCORE	GCCGCC	pathogen/ ethylene/ jasmonate response	1	0	0	1	1	0.4
HBOXCONSENSUSPVCHS	CCTACCNNNNNNNCT	H-box; light/ defence response	1	0	0	0	0	0.0
IRO2OS	CACGTGG	G-box; Fe-deficiency response	1	0	0	0	0	0.0
LTRE1HVBLT49	CCGAAA	low temperature response	0	0	0	0	1	0.1
LTRECOREATCOR15	CCGAC	low temperature response	0	1	0	2	4	0.5
MYB1AT	WAACCA	dehydration response	1	0	0	0	0	0.6
MYB2CONSENSUSAT	YAACKG	general stress/ dehydration response	2	1	0	0	1	0.8
MYBCORE	CNGTTR	dehydration response	3	1	2	1	2	1.4
MYCATERD1	CATGTG	dehydration response	1	0	0	0	0	0.8
MYCATRD22	CACATG	dehydration response	1	0	0	0	0	0.1
MYCCONSENSUSAT	CANNTG	temperature/ dehydration response	11	1	1	2	2	6.1
RYREPEATBNNAPA	CATGCA	Seed-specific expression	2	0	0	0	0	0.4
SEBFCONSSTPR10A	YTGTCWC	pathogen response	0	0	0	0	2	0.4
SITEIIAOSPCNA	TGGGCCCGT	meristematic tissue-specific expression	0	1	0	0	0	0.0
SURECOREATSULTR11	GAGAC	sulphur response	2	0	0	0	0	1.0
T/GBOXATPIN2	AACGTG	jasmonate response	2	0	0	0	0	0.0
WBOXATNPR1	TTGAC	pathogen response	0	1	0	0	3	1.6
WBOXNTERF3	TGACY	wound response	5	0	0	1	2	2.9
WRKY71OS	TGAC	pathogen response	13	1	0	2	4	4.8
		TOTAL:	68	10	5	24	35	27

The distribution of stress-associated CREs and predicted transcription start sites (TSS) within the 5'LTR and UTR of the five retrotransposon families is shown in Figure 3.14. No predicted TSS was found for *Edel*, and the TSS of *Tvv1* was localised to 377bp within the UTR.

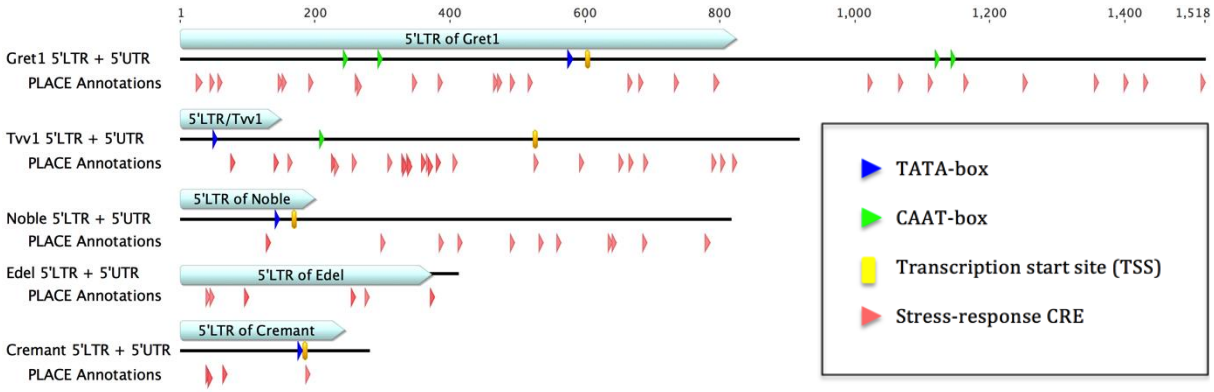


Figure 3.14 Distribution of stress-associated *cis*-regulatory elements (CREs) in five retrotransposon families. The 5' end of each retrotransposon is shown, beginning at the 5'LTR and ending immediately before the first codon of the internal coding regions. All CRE motifs from the PLACE database associated with stress response are indicated by red arrows.

3.4.4 Retrotransposons in grapevine transcript databases

Several of the retrotransposon families are highly represented in *V. vinifera* transcript databases. There was no correlation between the number of complete elements in the grapevine genome for a given family and the representation of this element in EST databases (Figure 3.15 and Figure 3.16). The EST hits returned were almost exclusively from transcript databases generated from stressed *V. vinifera* tissue samples.

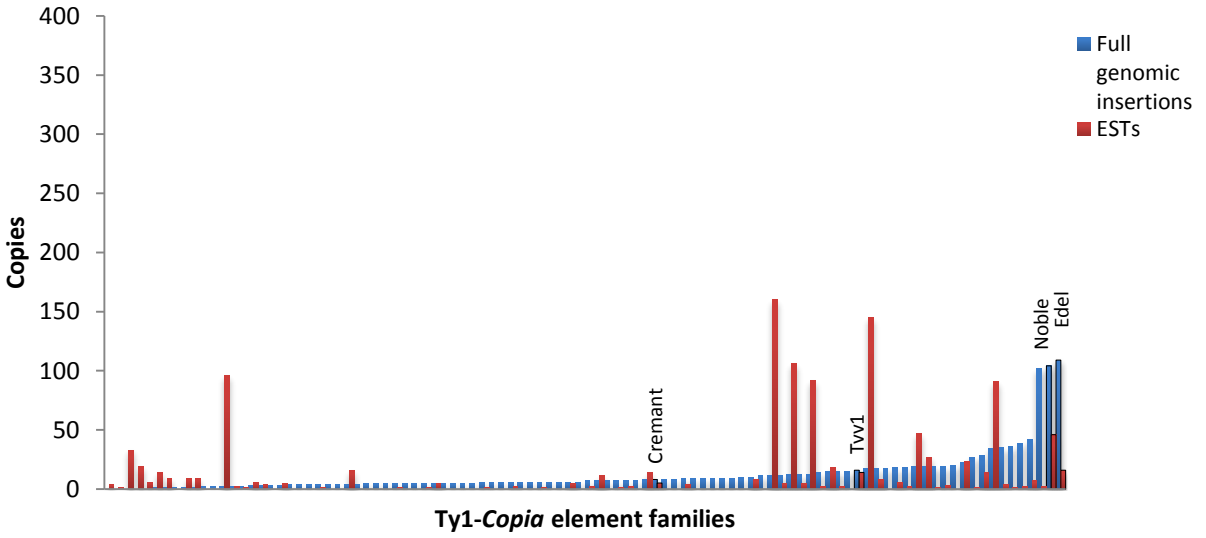


Figure 3.15 Comparison of genomic copies and EST abundance for Ty1-*Copia* elements in grapevine. Only complete copies are counted. EST counts indicate the number of matching sequences in the NCBI Genbank EST database. Families studied in further detail in the following sections are named.

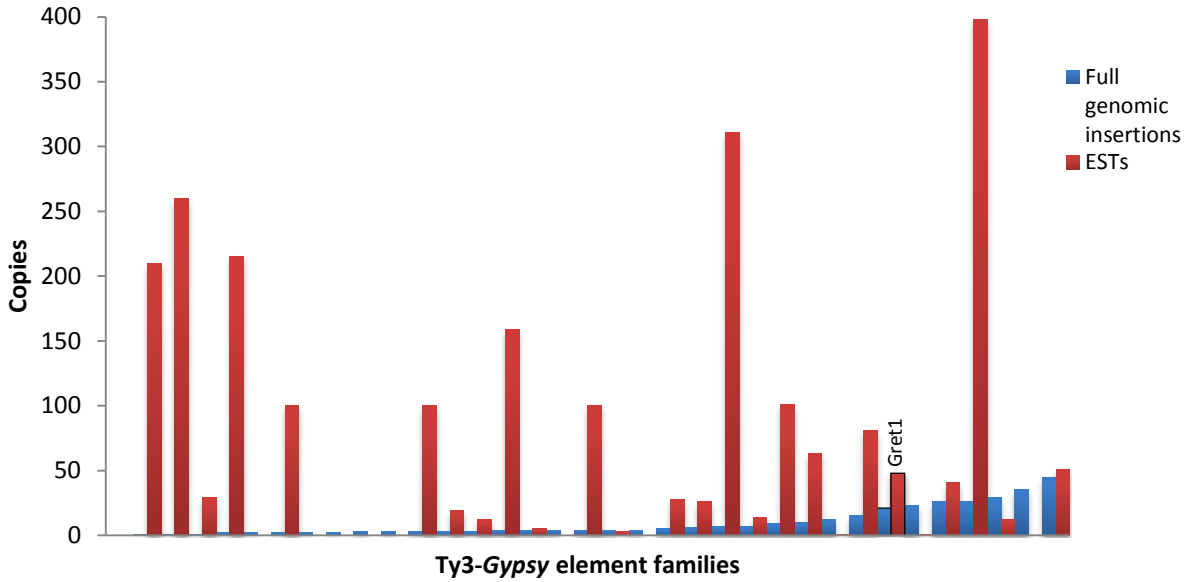


Figure 3.16 Comparison of genomic number and EST abundance for Ty3-*Gypsy* elements in grapevine. Only complete copies are counted. EST counts indicate the number of matching sequences in the NCBI Genbank EST database. Families studied in further detail in the following sections are named.

3.5 Discussion

The term 'mobilome' is frequently used to describe the portion of a genome that is associated with mobilisation events. The mobilome is by its nature dynamic, with transposition adding new elements that are paralysed and gradually eroded by mutations, epigenetic modifications and recombination events. By studying repetitive elements within whole genome datasets, we are able to analyse a snapshot of this dynamic system. From comparisons of the abundance and integrity of sequences from different TE families inferences can be made about the current and historical activities of these families.

3.5.1 Insertion numbers

The grapevine retrotransposon families represented in the Repbase Update database vary greatly with respect to their contribution to the total genome. For most retrotransposon families, fewer than 10 insertions exist. In contrast, the largest nine Ty1-*Copia* and seven Ty3-*Gypsy* families account for more than half of the genomic insertions in each of their respective superfamilies. The abundance of these families suggests high rates of activity in the past (transposition bursts) or the evasion of host silencing mechanisms for long periods of time.

In particular, sequence fragments from a few retrotransposon families (6 Ty1-*Copia* and 8 Ty3-*Gypsy*) are present in far higher number than those of other families. It is possible that proliferation of these elements may have primarily occurred prior to the polyploidisation of grapevine. The low proportion of intact retrotransposons within these families, an indication that most insertions are relatively ancient, further supports this hypothesis. If this is true, then it is probable that other element families, which are present in lower insertion numbers but show similarly high degradation of individual elements, also underwent most of their more limited transpositional activity in a diploid ancestor of modern grapevine.

3.5.2 Retrotransposon lineages

The grapevine genome contains TE insertions representing each of the retrotransposon clades described in plants (Figure 3.9). Definition of the Ty1-*Copia* superfamily according to clades is still considered a work in progress, however. The indistinct nature of this endeavour is reflected by the clustering of nineteen grapevine retrotransposons into a single branch which includes the eponymous elements of the *Maximus* / *Sire* and *Oryco* clades. Thirty-seven of the grapevine retrotransposon families also form a branch that is quite distinct from any of the previously defined retrotransposon clades.

The high copy number of certain retrotransposons within the grapevine genome is consistent with the current understanding of the replicative nature of class I TEs and the state of most other plant genomes studied to date (see 2.2.2). However, the retrotransposons within the grapevine genome can be sorted into distinct and diverse retrotransposon families (136 currently defined). Three hypotheses could be used to explain the presence of such a large number of retrotransposon families within the genome.

1. A large number of retrotransposon families existed in species ancestral to *V. vinifera* and have been directly inherited in the modern grapevine genome.

2. A smaller number of ancestral retrotransposon families were inherited by grapevine from an ancestral genome. These elements have multiplied and diversified within the grapevine genome.

3. Horizontal transfer of retrotransposons between species has led to the infection of the grapevine genome with new retrotransposon families since the speciation of *V. vinifera*.

The first hypothesis is predicted to produce a genome with high numbers of degraded retrotransposon copies, each eroded by generations of mutation and recombination events. This appears to be the case for many of the retrotransposon families of modern grapevine, particularly those that are the most abundant. However, this hypothesis is insufficient to completely describe the current situation of the mobilome. Several hundred copies of complete retrotransposon elements exist (Figure 3.4 and Figure 3.5) and serve as evidence of the recent mobilisation and amplification of these families.

In the case of the second hypothesis, the relatedness of retrotransposon families within the grapevine genome would be generally higher than relatedness of retrotransposon families from different species. Phylogenetic comparisons of grapevine elements with retrotransposons from the genomes of other angiosperms (Figure 3.11 and Figure 3.12) would be expected to produce a few distinct grapevine-specific clades representing the limited lineages from which modern retrotransposons families have descended. This pattern is seen for *Arabidopsis thaliana*, particularly within the Ty1-*Copia* element superfamily, and for *Glycine max*, particularly in the Ty3-*Gypsy* superfamily. However, the retrotransposon families of grapevine and other angiosperms included in the analysis show less inter-species definition, with some elements grouping into clades defined by species boundaries and others more closely related to elements from other species.

With regards to the third hypothesis, horizontal transfer of elements between species would result in retrotransposon families with highest sequence similarity to elements in foreign genomes. The

distribution of *V. vinifera* retroelements among repeat elements from other angiosperm species in the inter-specific cladograms does suggest that certain families may not have followed a pattern of linear inheritance. In particular, several families appear to fall into well-defined clades that have no other members in grapevine (Figure 3.12). Evidence for transfer of retrotransposons between plant species has also been found in *Arabidopsis* (PEREIRA 2004) as well as rice, soybean, maize and sorghum (ROULIN *et al.* 2009; DU *et al.* 2010).

3.5.3 Retrotransposon expression

According to the current understanding of the replication of class I TEs, transcription of each element is expected to begin within the R-region of the 5' LTR. Computational transcription start site prediction identified motifs within the sequences of *Gret1*, *Cremant* and *Noble* at the expected site within the 5'LTR. However, no TSS could be predicted for *Edel*, and the TSS of *Tvv1* was localised to 377bp within the UTR. This highlights the limitations of using plant transcription factor binding site databases to characterise the promoter sequences of retrotransposons, which are more structurally similar to retroviruses than to plant genes. The 5' boundaries of retrotransposon transcripts need to be experimentally determined in order to define these with confidence.

The LTRs of the five plant retrotransposon families studied in further depth contain multiple stress-related CREs. In addition to the information that this provides about conditions under which these TEs are potentially transcribed, it is also worth considering whether genes directly downstream of these elements may be transcriptionally co-regulated by retrotransposon 3' LTRs. For examples, the LTRs of *Gret1* contain an abundance of stress-related CREs. As explained (see 2.2.10) it is the insertion of a *Gret1* retrotransposon in the promoter of the VvMYBA1 transcription factor that prevents synthesis of red anthocyanin pigments in white grape varieties.

The abundance of retrotransposon sequence fragments in EST databases indicates that several grapevine retrotransposons can become transcriptionally active. These databases represent the sum of multiple tissue types across different grapevine varieties. The great majority of the results with sequence similarity to retrotransposon sequences were derived from grapevine tissues subjected to stress treatments. While EST hits are not a measure of relative expression levels, they do give an indication of which retrotransposon families can become transcriptionally active, and provide information about the tissue types and experimental conditions under which their expression has been observed.

3.6 Conclusions

A large proportion of the grapevine genome is composed of complete and partial sequences of grapevine retrotransposons. The relative integrity of these elements suggests that while certain families are ancient, and probably predate the speciation of grapevine, other families have undergone recent transposition events and may still escape host silencing under certain conditions. Analysis of conserved motifs within retrotransposon sequences and the representation of these elements in EST databases reveal that most grapevine retrotransposons are transcriptionally active when the plant is exposed to environmental stresses.

As the cost of large-scale sequencing continues to fall, we can expect collections of resequenced genomes and transcriptomes to expand. Though our understanding of the mobilome is less mature than that of the exome, bioinformatics tools which enable *in-silico* characterisation of retrotransposon families and comparative genomic approaches will increasingly become the first call for researchers seeking to characterise the impact of mobile DNA.

Chapter 4

Retrotransposon-based Markers for Genotyping Grapevines

4.1 Overview

Molecular genotyping is widely used for the discrimination and classification of species, populations and individuals. The identification of somatically propagated agricultural crops by DNA profiling depends on characterisation of genomic regions that are mutable within somatic tissues. With their dynamic nature and ubiquitous presence, transposons are particularly well suited to development as molecular markers. The repetitive insertion of retrotransposons results in stable, dominant DNA polymorphisms often at rates above that of SNP mutagenesis. The high copy number of retrotransposon families in eukaryotic genomes also allows numerous markers to be scored simultaneously in a single PCR reaction. Although researchers have attempted to utilise retrotransposon-based markers for the discrimination of grapevine varieties and clones, their success has been limited by the lack of sequence data available for the mobilome as a whole and the resolution of polyacrylamide gel-based displays.

Using results obtained from *in silico* mining of retrotransposon sequences in the publically available draft grapevine genome sequence (see Chapter 3), a sequence-specific amplified polymorphism (S-SAP) approach was used to amplify DNA sequences flanking four specific retrotransposon families. By using an ABI capillary sequencer to detect fluorescently labelled primer tags, transposon display profiles were generated that permit comparison of several hundred markers simultaneously at single base-pair resolution and with much higher throughput than gel-based analyses. Polymorphic markers were then isolated by replacing fluorescent tags with biotin and separating the amplification products by gel electrophoresis. The sequence data of these individual fragments was used to develop locus-specific inter-retrotransposon binding polymorphism (IRBP) primer sets that proved useful for direct genotyping of plant material preserved on FTA cards (Whatman).

This protocol demonstrates the use of whole-genome sequence data to produce an effective transposon-based molecular marker system. Marker profiles produced using a single adapter primer and any of three retrotransposon families (*Gret1*, *Edel* and *Noble*) allowed differentiation among all of the 14 *Vitis* varieties tested except for those of the Pinot varieties. Inter-clonal polymorphism was observed within Pinot noir, Chardonnay and Sauvignon blanc for these three retrotransposon families.

4.2 Introduction

Since the invention of DNA profiling (JEFFREYS *et al.* 1985), the ability to identify and characterise individuals at a molecular level from a small tissue sample has become a cornerstone of modern forensics, medical diagnosis and genetics. Due to the vast size of individual genomes, the speed and affordability of DNA testing depends on the specific examination of only the most variable regions of the genome.

4.2.1 Molecular markers for genotyping

With the advent of PCR a variety of molecular techniques have arisen to rapidly differentiate organisms by genotype. These marker systems can be generally grouped into two types. In the case of the first type, a pair of oligonucleotide primers is used to amplify a specific locus in the genome known to be a site of common sequence variation. Examples of such systems include sequenced characterized amplified region (SCAR) markers (PARAN AND MICHELMORE 1993), which detect single-nucleotide polymorphism (SNP) variation, and simple sequence length polymorphism (SSLP) markers (TAUTZ 1989), which detect variance in 1-6bp tandem repeat counts. Genotype scoring with these marker types generally involves PCR and gel electrophoresis, rudimentary procedures for most molecular biology laboratories. However, the design of the primer sets upon which these techniques are based relies on precise sequence information of the target locus.

Marker systems of the second type make use of primers that anneal to degenerate or highly repetitive sequences in the genome. For the development of these markers general information about genomic sequences, such as common motifs or repeat types, can be substituted for specific knowledge about individual loci. In addition, the non-unique nature of these sequences results in the amplification of multiple loci simultaneously, producing a complex 'DNA fingerprint' pattern. As with the first marker type, genotyping can be based on SNP variation, as is the case with the random amplification of polymorphic DNA (RAPD) technique (WELSH AND MCCLELLAND 1990; WILLIAMS *et al.* 1990), or variations in microsatellite repeat counts, in the case of inter-simple sequence repeat (ISSR) PCR (ZIETKIEWICZ *et al.* 1994).

Since these techniques rely on the exponential amplification of multiple targets in a single reaction, slight variations in DNA purity and reaction conditions can alter the results. Consequently, these DNA profiles often suffer from lack of reproducibility. The amplified fragment length polymorphism (AFLP) technique (VOS *et al.* 1995) went some way to improve this situation, by appending oligonucleotide adapters to enzyme-fragmented DNA before amplification. Primers are used which bind the synthetic adapters, rather than the genomic DNA, thereby avoiding DNA impurities and modifications that may hinder primer annealing.

4.2.2 Transposon profiling

As recognition of the abundance and dynamic nature of transposable elements (TEs) has increased, these elements have begun to attract attention as candidates for molecular marker development (Figure 4.1). Waugh and colleagues (1997) developed the sequence-specific amplified polymorphism (S-SAP) technique, which incorporates transposon-specific primers into an AFLP-based system. Markers produced by this technique have since been found to be more informative than SNP or SSR-based markers in multiple agricultural crop species, including oats (YU AND WISE 2000), tomato and pepper (TAM *et al.* 2005), sweet potato (BERENYI *et al.* 2002), pea (ELLIS *et al.* 1998) and grapevine (LABRA *et al.* 2004).

Profiling of transposon insertions by S-SAP begins with the enzymatic digestion of purified genomic DNA, followed by the ligation of short adapter sequences. In addition to the reduction in throughput caused by these additional reactions, there is a possibility that epigenetic modifications of the DNA may differentially influence enzymatic digestion, thereby contributing towards marker polymorphism. To address these issues two PCR-based techniques that directly target retrotransposons in genomic DNA have been proposed. The first, inter-retrotransposon amplified polymorphism (IRAP), involves amplification of the stretches of DNA between retrotransposon insertions. The second, retrotransposon-microsatellite amplified polymorphism (REMAP), produces amplicons corresponding to regions between SSRs and adjacent retrotransposons (KALENDAR *et al.* 1999). Both techniques have been used to genotype several plant species including barley, clementine, wheat, apple, grapevine, banana, pea and sawgrass (For reviews see (KUMAR AND HIROCHIKA 2001; KALENDAR AND SCHULMAN 2007).

A further retrotransposon-based marker system, retrotransposon-based insertional polymorphism (RBIP), uses primers with specificity to known transposon insertion sites to test for the presence or absence of a particular insertion (FLAVELL *et al.* 1998). While this technique requires sequence data for the polymorphic site, screening involves only a standard PCR, which is resolved on an agarose gel. Since primers can be designed to amplify both the presence and the absence of a transposon insertion, this system produces co-dominant marker information.

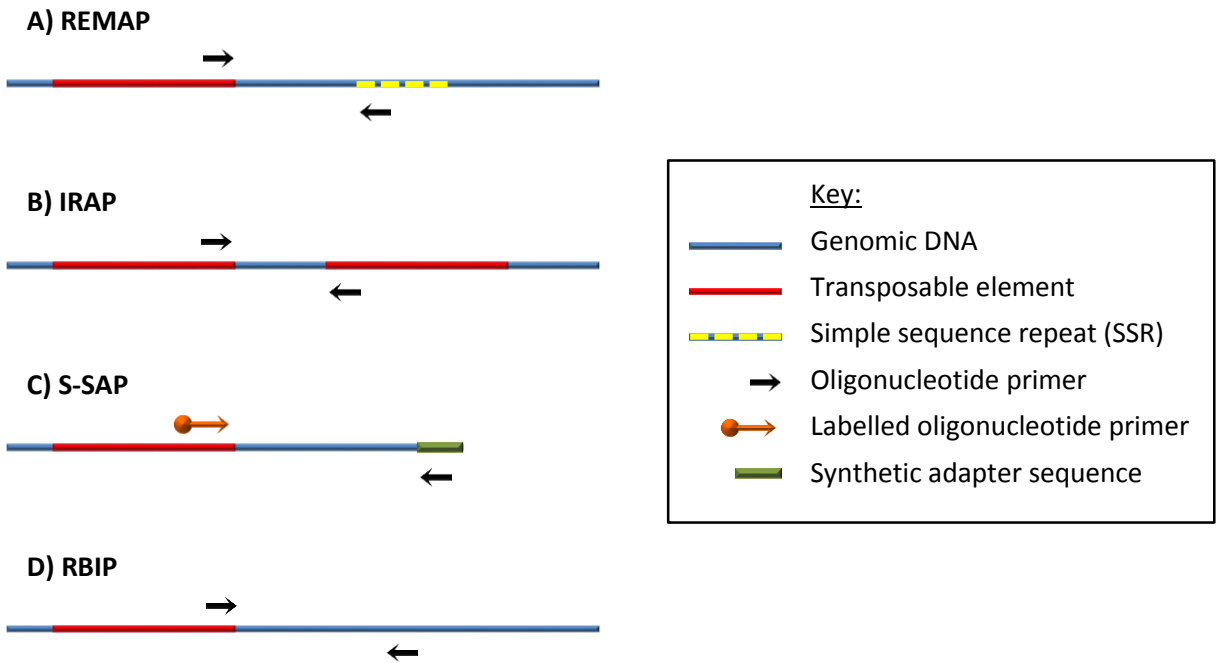


Figure 4.1 Molecular marker systems that make use of TE polymorphism. A) Retrotransposon-microsatellite amplified polymorphism (REMAP) involves the simultaneous amplification of multiple loci between elements of a retrotransposon family and adjacent simple sequence repeats. **B)** Inter-retrotransposon amplified polymorphism (IRAP) involves the simultaneous amplification of DNA regions between retrotransposons. **C)** Sequence-specific amplified polymorphism (S-SAP) involves the simultaneous amplification of multiple loci between elements of a retrotransposon family and adapters ligated at restriction enzyme cut sites. **D)** Retrotransposon-based insertional polymorphism (RBIP) primer pairs amplify single loci between a retrotransposon and a specific genomic site.

4.2.3 The use of retrotransposons as markers in grapevine

Considering the value and size of the grape industry, it is unsurprising that almost every molecular marker type has been applied to the grapevine genome in some form. To date, most researchers have used sequence-specific SSR markers for identification of quantitative trait loci (QTL), parentage studies, and cultivar identification (Bowers *et al.* 1999; Regner *et al.* 2000; Doligez *et al.* 2002; This *et al.* 2004). More recently, the applicability of transposon display techniques for inter- and intra-variety identification has been tested by several researchers.

As discussed (see section 2.2.9), two Ty1-*Copia* elements, *Vine1* (Verries *et al.* 2000) and *Tvv1* (Pelsy and Merdinoglu 2002), and one Ty3-*Gypsy* element, *Gret1* (Kobayashi *et al.* 2004), were described prior to the completion of the *V. vinifera* genome sequencing projects. Initial transposon display experiments used these elements to distinguish between different *Vitis* species and among *V. vinifera* varieties, finding that they produced markers with a higher average polymorphism than AFLP markers (Pelsy *et al.* 2003; Labra *et al.* 2004; D'Onofrio *et al.* 2010). To date, retrotransposon-based marker have demonstrated intra-varietal polymorphism for clones of the cultivars Traminer, Malbec and Syrah, but have not proved useful for distinguishing clones of the Sangiovese, Niellucciu,

and Prugnolo and Pinot varieties (LABRA *et al.* 2004; D'ONOFRIO *et al.* 2009; STAJNER *et al.* 2009; D'ONOFRIO *et al.* 2010).

Following the public release of the assembled grapevine genome, Moisy and colleagues (2008) identified 10 novel Ty1-*Copia* retrotransposon families from degenerate PCR amplification of RNase H domains and comparison of the cloned products with genome sequence data. Eight of these were tested with S-SAP on 10 *Vitis* accessions (including 7 *vinifera* varieties). Each of the retrotransposon families produced a high proportion (>50%) of markers that were polymorphic between varieties. No inter-clonal comparisons were reported.

While each band scored by these marker techniques is typically assumed to be representative of a retrotransposon insertion, this is not necessarily the case. Other mutation types, including SNPs, can affect both the restriction digestion sites and primer specificity. The contribution of such mutations to S-SAP profiles is unknown, since amplification products have rarely been sequenced and mapped back to the grapevine genome. Wegschieder and colleagues (2009) were able to differentiate between certain clones of Pinot cultivars, using markers generated by S-SAP amplification with universal degenerate plant retrotransposon primers. However, only four of the polymorphic marker bands were sequenced, and just one of those could be associated with a retrotransposon. In addition, all four polymorphic bands were a result of restriction site polymorphisms, suggesting that the universal primers were serving to anchor an SNP-based analysis, rather than tag polymorphic TE insertions.

Both the IRAP and REMAP transposon display techniques have also been tested for grapevine genotyping. No polymorphisms between clones have been detected with IRAP; in fact only one primer has shown markers that differentiate between varieties. The application of REMAP proved successful for cultivar identification, but displayed limited success with clonal identification, capable of distinguishing a total of four clones to date. Moreover, the majority of REMAP polymorphisms have proved to be a result of ISSR polymorphisms rather than retrotransposon mobility (CARCAMO *et al.* 2010; CASTRO *et al.* 2012).

It is clear therefore that while the mutagenic capacity of transposable elements makes them good candidates for marker design, the identification of suitable target elements and optimised primer design are essential to the realisation of their potential. Furthermore, care must be taken when drawing conclusions regarding retrotransposon activity from polymorphism patterns. Without locus-specific sequence data, individual polymorphisms cannot be confidently associated with retrotransposon activity.

4.2.4 Improved transposon-based genotyping in grapevine

In this chapter, the reproducibility and information content of transposon display profiles generated using the techniques described above are compared using genomic DNA from several grapevine varieties and clones. Since IRAP and REMAP are both fairly simple and rapid techniques, the reproducibility and resolution of profiles generated by these techniques was first tested with primers targeting five grapevine retrotransposon families (*Gret1*, *Edel*, *Cremant*, *Noble* and *Tvv1*). In comparison with these methods, the generation of consistent, high-resolution S-SAP profiles from grapevine DNA samples is described. The protocol used is based on that of Syed and Flavell (2007), with modifications to primer design, the inclusion of fluorescent primer labels in place of ^{33}P radioactive tags and the use of computer software to detect bands (Figure 4.2). For these experiments, different fluorescent dye chemicals were used to label primers targeting each of four of the selected retrotransposon families (*Gret1*, *Edel*, *Cremant* and *Noble*) so that their amplification products could be simultaneously scored by capillary electrophoresis.

Restriction enzymes with 4bp (*Mse*I) and 6bp (*Eco*RI) recognition sites were used for co-digestion of DNA, to allow selective amplification with different adapter primers from the same preamplified template, and so that AFLP profiles to be generated from the same preamplified samples. The procedure benefitted from the optimisation of several variables, while the alteration of others had little or no effect on the S-SAP profiles within the parameter ranges tested. As expected, the marker density and polymorphism of the S-SAP profiles is dependant on the retrotransposon family targeted during primer design. Individual polymorphic bands recovered by the purification of biotin-tagged fragments enabled the design of locus-specific RBIP primers. These primers were then used to rapidly genotype plant material.

The activity of certain restriction enzymes is hindered by the methylation of cytosine bases at the recognition site. As a result, S-SAP profiles produced from DNA digested with these enzymes reflect both epigenetic and genetic variation. This derivative technique has been termed methyl-sensitive transposon display (M-STD; (ZERJAL *et al.* 2009). To confirm that this tool can be used to display variation across grape epigenomes, M-STD profiles were produced by parallel processing of samples with two isoschizomer enzymes. The enzymes *Msp*I and *Hpa*II were used, both of which digest unmethylated DNA at the CCGG recognition sequence. However, *Msp*I digestion is blocked by methylation of an external cytosine on at least one strand, while *Hpa*II digestion is blocked by complete methylation (both strands) of either cytosine.

Finally, plant tissue stored on FTA cards (Whatman) was tested as a template for S-SAP and RBIP reactions in place of purified DNA. These cards are impregnated with a proprietary solution designed to enable small card punches to be used as templates for DNA amplification reactions after a few

wash steps. The convenient collection and storage of samples in this form, together with the elimination of DNA extraction steps could improve throughput when genotyping multiple samples.

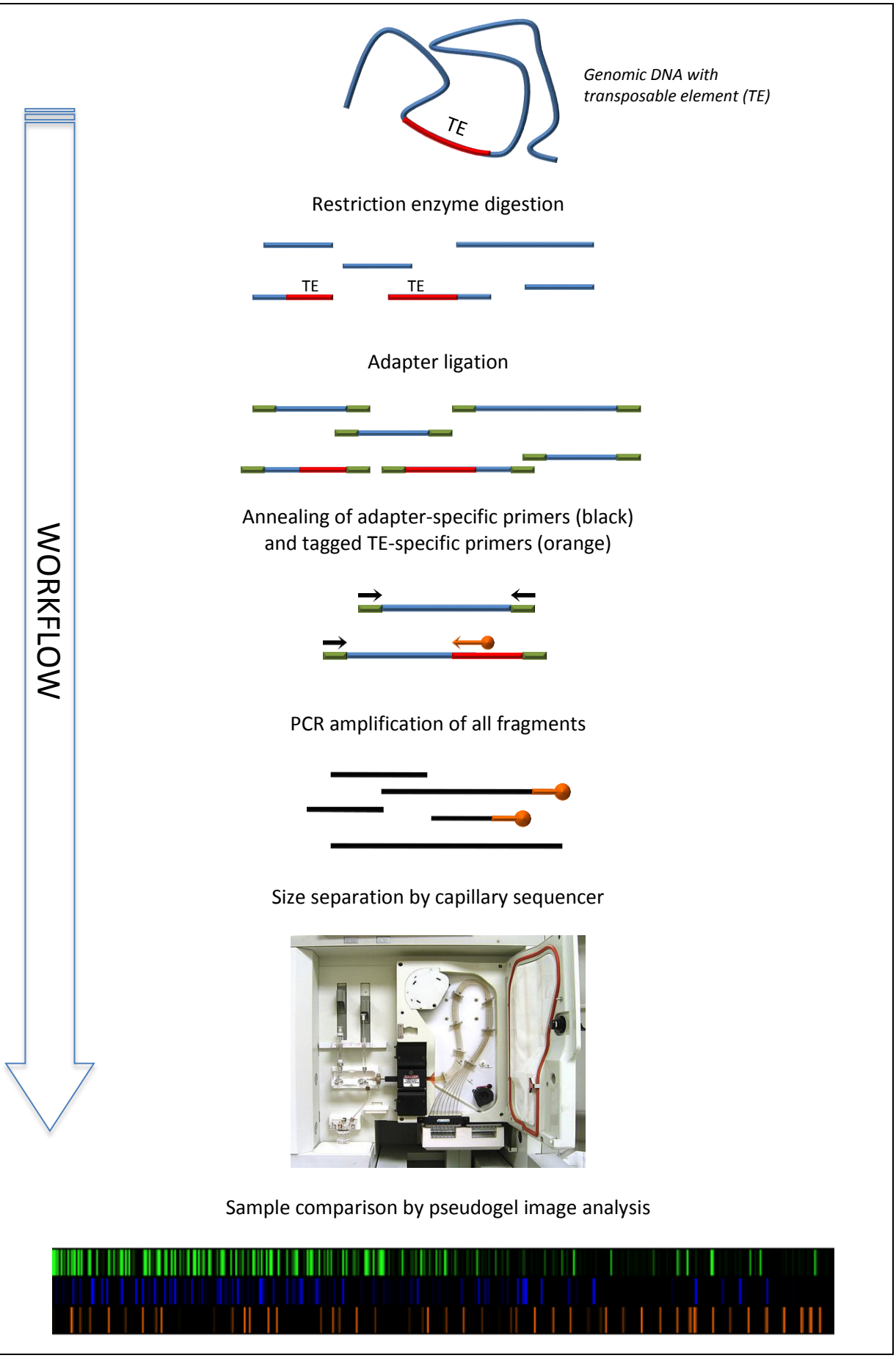


Figure 4.2 Workflow of a sequence-specific amplified polymorphism (SSAP) experiment using fluorescently labelled probes.

4.3 Methods

4.3.1 Sample preparation

Thirty-two grapevine genotypes from the Lincoln University grapevine germplasm collection were selected as a sample set. Among these were 12 different *V. vinifera* varieties, including 16 clones from Pinot varieties and 3 clones each of Sauvignon blanc and Chardonnay. The rootstocks R110 (Richter 110; *V. berlandieri* x *V. rupestris*) and S04 (Selection Oppenheim Nr. 4; *V. berlandieri* x *V. riparia*) were also included. Total genomic DNA was extracted from young leaf material using the NucleoSpin Plant II kit (Machery-Nagel) according to the manufacturer's instructions, except that the lysis buffer was supplemented with 5% (w/v) PVPP and 1% (v/v) 2ME. To verify technical reproducibility, duplicate samples of 7 vines (5 Pinot noir clones, Riesling and Gewürztraminer) were collected and processed separately. Extractions were quantified using a Qubit fluorometer (Life Technologies) and purity was assayed with a NanoDrop spectrometer (Thermo Fischer Scientific). The integrity of extracted DNA was assayed by running 100ng of each sample on a 1% (w/v) TBE agarose gel at 100V for 20 minutes.

4.3.2 Primer design

For each retrotransposon family to be tested, a recent element insertion (indicated by high LTR identity and complete open reading frame) was selected as a template for primer design by searching for elements with perfect LTR identity and a complete internal ORF. Primers for amplifying DNA flanking retrotransposon insertion sites were designed with specificity to the extreme 5' end of the LTRs (Gret1 LTR rev(0), Edel LTR rev(0), Cremant LTR rev(0), Noble LTR rev(0), Tvv1 LTR rev(0)) or to the extreme 3' ends (Gret1 LTR for(0), Edel LTR for(0), Cremant LTR for(0), Noble LTR for(0), Tvv1 LTR for(0)), so that selective bases could be added to the primers to exclude amplification of TE internal regions, as recommended by Syed & Flavell (2007). All primers were designed to have a T_m of approximately 60-65°C (SANTALUCIA 1998). Primers designed to anneal to microsatellite repeat regions for REMAP contained 9 dinucleotide repeats and one selective base at the 3' end (Microsat-GA(C) and Microsat-CT(G)). All primer sequences are given in Appendix C.

To maximise the number of elements amplified by S-SAP, the BLAT program (KENT 2002) was used to extract all LTRs copies from the 12X PN40024 grapevine genome with greater than 90% identity to the Repbase ancestral LTR sequences. Recovered sequences with greater than 50% coverage of the parent element were aligned using MUSCLE (EDGAR 2004). Downstream-oriented primers were then designed with specificity to highly conserved regions within 100bp of the 3' end of the LTRs, with a T_m of approximately 60-65°C.

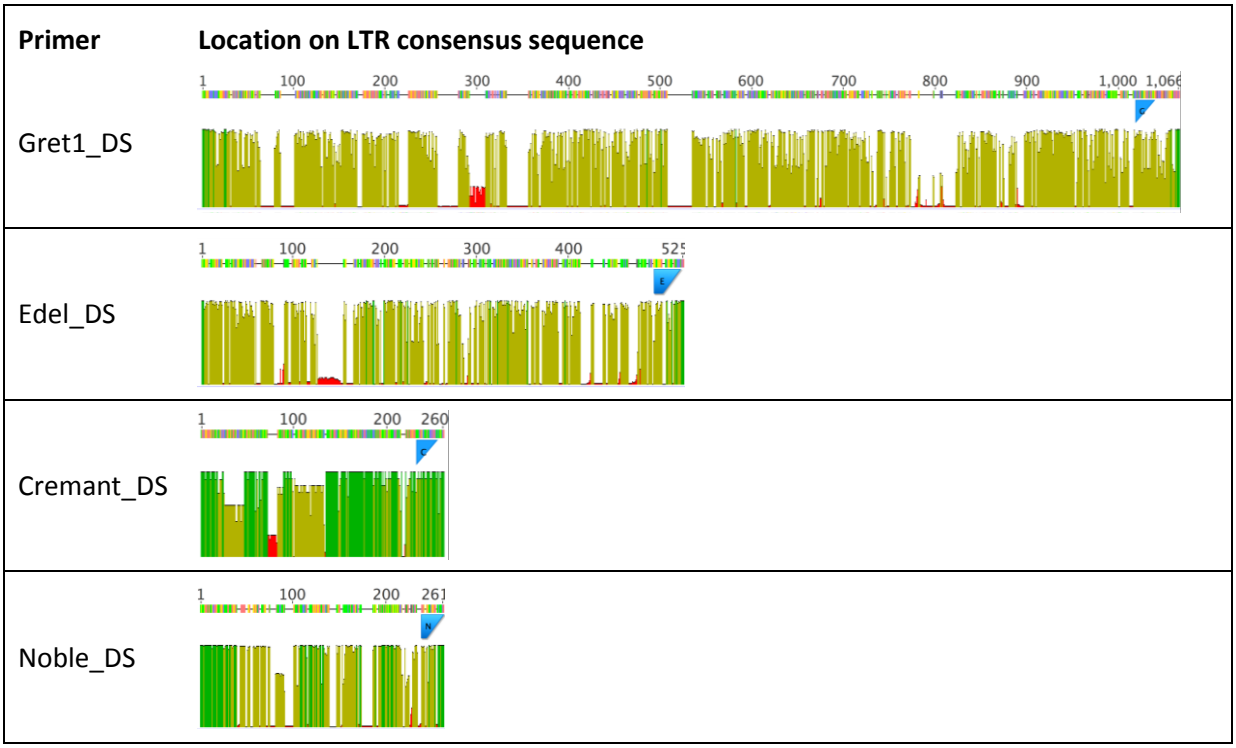


Figure 4.3 Downstream S-SAP primer binding sites. Binding sites for downstream-orientated primers are indicated on the TE LTR consensus sequences by blue arrows. Graphs show the proportion of all LTR sequences within the TE family matching the consensus sequences (per cent identity of the consensus) at each base. Green = 100% identity, Yellow = 50% identity or above, red = below 50% identity.

Fluorescent dye modifications were added to the 5' termini of LTR-specific primers. Dye chemicals with minimal excitation peak overlaps (ABI DS-33 dye set) were selected to allow multiplex fragment analysis (Table 4.1). For AFLP analyses, a VIC-labelled *Eco*RI adapter primer was used.

Table 4.1 Fluorescent dye modifications on S-SAP primers.

Retrotransposon family	Dye	False colour	Supplier
Gret1	FAM	Blue	IDT
Edel	VIC	Green	ABI (Life Technologies)
Cremant	PET	Red	ABI (Life Technologies)
Noble	NED	Yellow	ABI (Life Technologies)
*Eco*RI adapter	VIC	Green	ABI (Life Technologies)

4.3.3 REMAP & IRAP

The protocols of Kalendar and Schulman (2007) were used for IRAP and REMAP experiments. The protocols for both amplification techniques are identical except for the primers that are used. In the case of REMAP reactions, one LTR-specific primer and one SSR-specific primer was used (see Appendix C for primer sequences). For IRAP, two LTR-specific primers were used. In both cases 20ng of genomic DNA was used as template in a 20µl reaction containing 1X PCR buffer (Qiagen), 0.2mM of each dNTP, 0.2mM of each primer and 1U of Taq polymerase (Qiagen). Thermocycling conditions were as follows: 94°C for 4 min, 40 cycles of 94°C for 40 s and 72°C for 2 min, followed by 72°C for 5 min.

Amplification products were checked for amplification on a 1% (w/v) agarose TBE gel stained with ethidium bromide (45 min, 85V). Products with high numbers of bands were subsequently run on a 20cm X 20cm 10% (w/v) polyacrylamide gel (run for 14hrs at 150V) and visualised by silver staining (Byun *et al.* 2009).

4.3.4 S-SAP with fluorescent dye-labelled primers

DNA restriction digestion

DNA digestion, adapter ligation and preamplification were based on the protocol of Syed & Flavell (2007) using the same primer sequences as these authors. Grapevine genomic DNA (500ng) was co-digested with 5 units each of *Mse*I and *Eco*RI (New England Biolabs) in 40µl reactions containing 0.1 mg/ml BSA and 1x NEB Buffer 4, incubated overnight at 37°C. Enzymes were denatured by incubating the reaction at 65°C for 20 minutes. To test the efficiency of co-digestion, genomic DNA samples of four Pinot noir clones (UCD5, UCD13, 113, and Mariafield) were also digested with both enzymes sequentially. Enzymes were removed by washing DNA with phenol/chloroform/isoamyl alcohol (25:24:1) precipitating in ethanol between digestions. These same samples were also co-digested for only 3 hours in separate reactions, to test whether artefacts arising from incomplete digestion or star activity contribute to the resulting S-SAP display profiles. Complete digestion was confirmed by running 5µl of each digestion reaction on a 1% (w/v) agarose TBE gel at 100V for 20 minutes.

Adapter ligation

Double-stranded adapters (Appendix E.1) were ligated to digested DNA in overnight reactions at 4°C using T4 DNA ligase (New England Biolabs). The 41µ reaction contained 200U T4 ligase, 1X T4 buffer, 1.25 mM *Mse*I adapter, 0.63 mM *Eco*RI adapter and the remaining 35µl of the DNA digestion reaction.

Preamplification

Adapter-ligated DNA fragments were preamplified in a 25µl reaction containing 2µl of adapter-ligated DNA, 1X PCR buffer (Qiagen), 0.2mM of each dNTP, 0.2mM each of the adapter-specific primers MseI(0) and EcoRI(0) (Appendix C) and 1U of Taq polymerase (Qiagen). Thermocycling conditions were as follows: 95°C for 1 min, 30 cycles of 94°C for 1 min, 60°C for 1 min and 72°C for 1 min, followed by 72°C for 7 min. Five microliters of each amplification reaction was checked on a 1% (w/v) agarose gel and the remainder diluted with 80µl TE buffer.

Selective amplification with fluorescent-tagged primers

Selective reactions were performed using one 5' fluorescently labelled retrotransposon LTR primer and one *Mse*I adapter primer with selective bases at the 3' end (Appendix C). To increase throughput and consistency, the preamplified DNA samples were diluted a further 10-fold and the reactions were prepared using an epMotion 5070 liquid handling robot (Eppendorf).

Reactions were performed in a 10µl volume with 4µl of the diluted preamplification products, 0.5µM each of one fluorescently labelled transposon primer and one MseI(XX) adapter-specific primer (where 'XX' represents 2 selective nucleotide bases), 2.5 mM $MgCl_2$, 0.25U Taq polymerase (Qiagen), 1X Taq buffer and 0.2mM of each dNTP. Touchdown thermocycling conditions involved an initial denaturation at 94°C for 1 min followed by 33 cycles of denaturation at 94°C for 30 s, annealing at 65°C for 30 s and elongation at 72°C for 30 s, with the annealing temperature decreasing by 0.7 °C after each cycle for the first 13 cycles.

Optimal reaction conditions were selected following comparison of results produced by varying $MgCl_2$ concentration (1.5mM, 2.5mM or 5.0mM), volume of (5-fold diluted) template DNA from the preamplification reaction (0.2µl, 0.4µl or 0.6µl), polymerase brand (Qiagen Taq Polymerase or Takara Ex Taq Polymerase), polymerase amount (0.25U or 1.25U) and reaction volume (10µl or 50µl) in separate experiments.

Pseudogel Analysis of Results

Products amplified using each of the four fluorescently tagged LTR primers were mixed in equal volumes for multiplex capillary electrophoresis. From the combined products, 0.8µl was added to 9.5µl Hi-Di Formamide and 0.5µl Genescan 1200 LIZ size standard (Applied Biosystems). Fragments were resolved using an ABI 3130xl capillary sequencer (Applied Biosystems, California, USA). Fragment peaks were size-calibrated and displayed as single-channel pseudo-gel images for comparison using Genographer V2.1.4 (http://sourceforge.net/projects/genographer/; (BENHAM *et al.* 1999).

4.3.5 Biotin S-SAP

To recover sequence data of specific bands, S-SAP amplification was repeated on preamplified samples according to the protocol above, except that the fluorescently labelled retrotransposon primer was replaced with a biotinylated primer of identical sequence. Amplification products were mixed with an equal volume of binding buffer (10mM Tris-HCl, 1mM EDTA, 2M NaCl, 0.1% v/v Tween 20) and 100ug pre-washed streptavidin-coated paramagnetic beads (Roche, Mannheim, Germany). After 15 minutes at room temperature with gentle agitation, the beads were washed twice with TE buffer and DNA was then eluted into 20µl of ultrapure water at 70°C as described by Holmberg and co-workers (2005).

Five microliters of eluted fragments were added to 15µl of denaturing loading dye (0.05% w/v bromophenol blue, 20mM ETDA, 95% v/v Formamide, 0.05% w/v xylene cyanol), heated at 95°C for 3 minutes and placed immediately on ice. Amplification products were separated by denaturing PAGE on a 12% (w/v) polyacrylamide, 8M Urea, 1X TBE gel at 55°C, 120V for 16hrs. Gels were visualized by silver staining (BYUN *et al.* 2009).

4.3.6 Band re-amplification

Bands of interest were identified by visual comparison with the pseudo-gel images and a polyacrylamide plug was taken using a 1ml pipette tip. Re-amplification was performed directly on the gel plug. The 50µl PCR reactions included 2µM unlabelled retrotransposon primer, 2µM MseI(0) primer (Appendix C), dNTPs (200µM each), 1.25U Ex Taq polymerase (Takara), and 1x Ex Taq buffer. Thermocycling conditions were as follows: 95°C for 1 minute; 30 cycles of 98°C for 10 sec, 55°C for 30sec, 72°C for 1 minute followed by 72°C for 5 minutes. Amplification specificity was verified by running 10µl of each reaction on a 1% (w/v) agarose TBE gel at 100V for 25 minutes. The PCR products were then cleaned with the AxyPrep PCR Clean-up Kit (Axygen) and sequenced using the MseI(0) primer (Appendix E.5).

4.3.7 RBIP marker design and testing

Band sequence data quality was checked using Geneious software (version 6; www.geneious.com) and screened using RepeatMasker (version 4.0; http://www.repeatmasker.org) to detect flanking sequences derived from repetitive elements. Genomic loci were determined by mapping sequences to the 12X PN40024 grape genome assembly using MegaBLAST (ZHANG *et al.* 2000).

Thirteen forward RBIP primers were designed on sequences that mapped to specific genomic loci using Primer3 (UNTERGASSER *et al.* 2012) with optimal T_m of 58°C. These locus-specific primers were tested in conjunction with the unlabelled transposon primer by PCR either on 20ng of purified

grapevine genomic DNA or on grapevine leaf material preserved directly on FTA cards (Whatman). The 20µl reactions included 1X PCR buffer (Qiagen), 0.2mM of each dNTP, 0.2mM of LTR rev(0) primer, 0.2mM of RBIP forward primer, 1U of Taq polymerase (Qiagen) and wither 20ng of genomic DNA or a single washed FTA disk. Thermocycling conditions were as follows: 95°C for 1 min, 40 cycles of 95°C for 30 s, 55°C for 30 s and 72°C for 1 min, followed by 72°C for 5 min.

4.3.8 Methyl-sensitive transposon display

Methyl-sensitive transposon display (M-STD) was performed according to the S-SAP protocol above, except that samples were digested with both *Msp*I and *Hpa*II restriction enzymes in separate parallel reactions. Accordingly, H/M oligonucleotide adapters and HpaII/MspI(0) primers (Appendix C) were used in ligation and preamplification reactions respectively. For selective amplification, the labelled LTR rev(0) primers were used, in conjunction with HpaII/MspI(CT) adapter-specific primer. Genomic DNA samples of two Pinot noir clones (UCD13 and 187) and Riesling were tested in duplicate with both the *Hpa*II and *Msp*I digestions.

4.3.9 FTA cards

To accelerate sample processing, grape leaf tissue preserved on Whatman FTA card was supplemented for purified DNA at two stages of the technique. In the first case, the restriction enzyme digestion, adapter ligation and preamplification steps of S-SAP were performed directly on washed 1mm punch disks according to FTA Protocol BC01 (WHATMAN). The diluted pre-amplification products were used for selective amplification as described above (see 4.3.4).

Leaf tissue collected on FTA cards was likewise used as a template for RBIP marker PCR reactions in place of purified DNA. For these reactions, a single 1.0mm punch was taken from cards pressed with grapevine leaf material. Each punch was washed with 70% (v/v) ethanol for 5 minutes before washing with FTA reagent according to the manufacturer's instructions. The PCR master mix was then added directly to the punch (**Error! Reference source not found.**).

4.4 Results

4.4.1 Sample preparation

DNA extractions from 100mg of young leaf tissue yielded 3.6µg (±1.7µg) of high quality (260/280 ratio between 1.75 and 1.85; 260/230 ratio > 1.90), intact DNA (Figure 4.4).

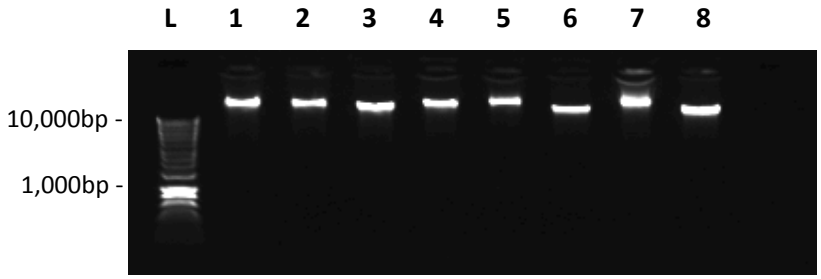


Figure 4.4 Visualisation of genomic DNA. Distinct bands indicate a high degree of integrity. One hundred nanograms of each sample was loaded on a 1% (w/v) agarose TBE gel and visualised after 20 minutes of running at 100V. Duplicate extractions were performed for each clone. L: Hyperladder I (Bioline), lanes 1-2: P. noir UCD5, lanes 3-4: P. noir UCD13, lanes 5-6: P. noir 113, lanes 7-8 P. noir Mariafield.

4.4.2 Direct amplification of genomic DNA with retrotransposon-specific primers

REMAP

Two SSR primers and eight LTR-specific primers were used in pairs to amplify Pinot noir UCD5 genomic DNA. No pairs resulted in more than six distinct bands (Figure 4.5).

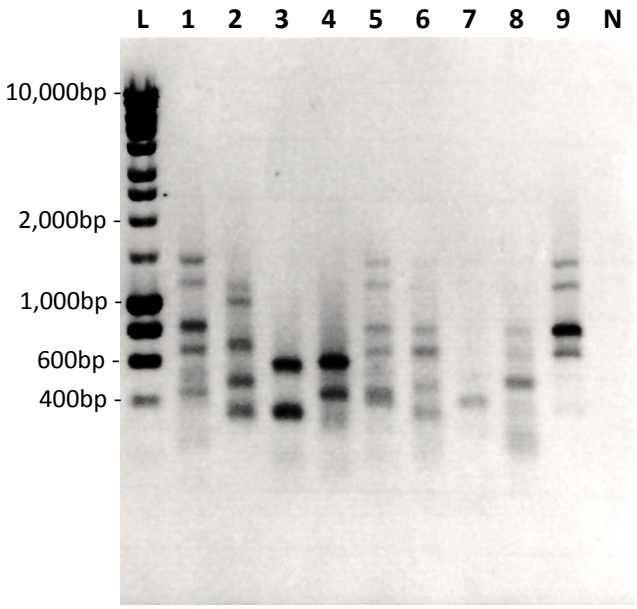


Figure 4.5 REMAP amplification products. Each amplification produced less than ten bands. Amplifications were performed on Pinot noir UCD5 genomic DNA using the Microsat-CT(G) primer and one LTR-specific primer as follows Lane 1: Edel for(0), 2: Edel rev(0), 3: Gret1 for(0), 4: Gret1

rev(0), 5: Tvv1 for(0), 6: Tvv1 rev(0), 7: Cremant for(0), 8: Cremant rev(0), 9: control – Microsat-CT(G) primer only, L: Hyperladder I (Bioline), N: no primer control. Amplification products were run on a 1% (w/v) agarose TBE gel for 45 minutes at 85V.

IRAP

The IRAP protocol did produce feature-rich banding patterns. Replicate reactions using the same template DNA produced consistent banding patterns. However, results were not consistent between separate DNA extractions from the same plant material (Figure 4.6). Changes to the T_m of the PCR reaction (ranging 50-70°C) or to the concentration of $MgCl_2$ (ranging 1.5mM-3.5mM) in the reaction mix did not improve reproducibility, nor did the inclusion of Q-solution (Qiagen) or DMSO as additive in the reaction mixture (data not shown).

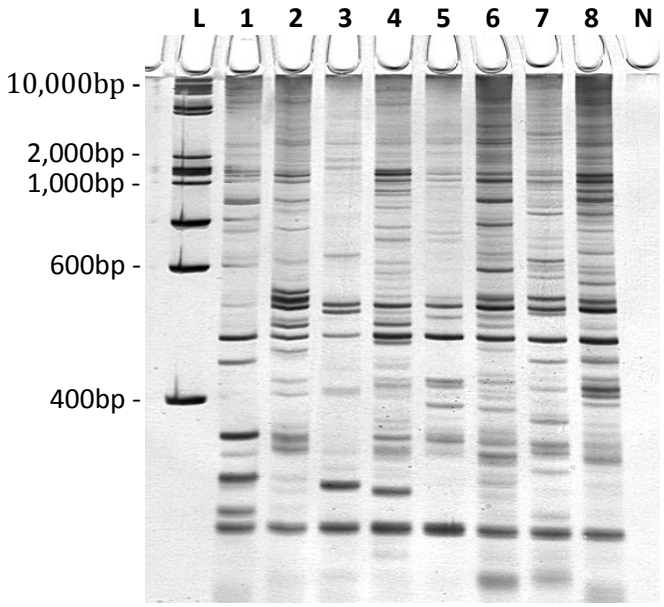


Figure 4.6 Inter-retrotransposon amplified polymorphism (IRAP) amplification products. Replicate samples processed separately (loaded in adjacent lanes) show unsatisfactory reproducibility. Amplification reactions included Cremant LTR for(0) and Gret1 LTR rev (0) retrotransposon primers. Key to lanes: L: Hyperladder I (Bioline), 1&2: P. noir UCD5, 3&4: P. noir UCD13, 5&6: P. noir 113, 7&8: P. noir Mariafield, N: no template control. Amplification products were run on a 10% (w/v) polyacrylamide gel for 14hrs at 150V.

4.4.3 S-SAP optimization

Sample processing

After overnight co-digestion with *Mse*I and *Eco*RI enzymes, gel electrophoresis showed complete fragmentation of DNA (Figure 4.7). The median fragment size is slightly over 100bp, which is expected since MseI has a 4bp recognition site sequence (TTAA; 4^4=128). Preamplification results showed even amplification of DNA fragments (Figure 4.8).

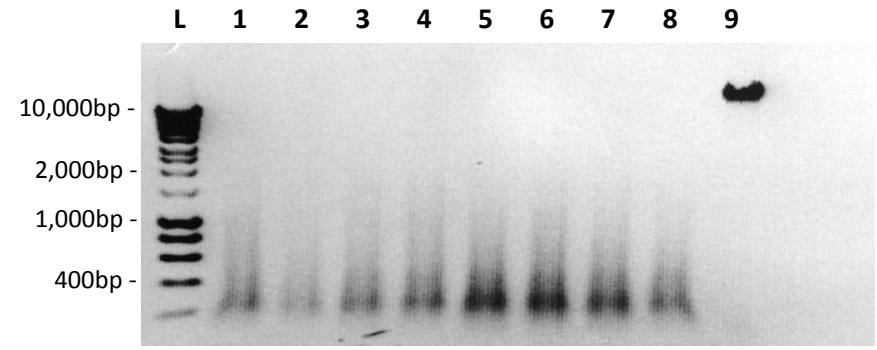


Figure 4.7 Restriction digest fragments from genomic DNA. Five microlitres of each digestion reaction (62.5ng) was visualised on a 1% (w/v) agarose TBE gel after electrophoresis at 100V for 30 minutes. Undigested genomic DNA is included (far right lane) for comparison. Key to lanes: L: Hyperladder I (Bioline), 1&2: P. noir UCD5, 3&4: P. noir UCD13, 5&6: P. noir 113, 7&8: P. noir Mariafield, 9: undigested P. noir UCD5 genomic DNA.

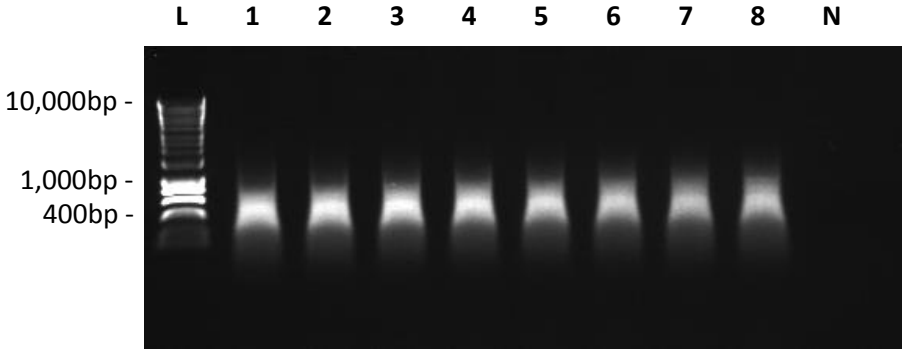


Figure 4.8 Visualisation of preamplified DNA fragments. Preamplification produces a library of fragments, most of which are between 200bp and 1kb. Five microlitres of preamplification reaction from duplicate sample processed separately were visualised on a 1% (w/v) agarose TBE gel after electrophoresis at 100V for 20 minutes. Key to lanes: L: Hyperladder I (Bioline), 1&2: P. noir UCD5, 3&4: P. noir UCD13, 5&6: P. noir 113, 7&8: P. noir Mariafield, N: no template control.

Fragment visualisation

Selective amplification with fluorescent dye-labelled primers produced fragments that could be clearly detected by capillary electrophoresis (Figure 4.9 A). Multiplexing products labelled with four different dyes was possible without compromising signal quality. Automated sizing of fragments was most reliable when sample peaks were of similar relative fluorescence intensity to those of the size standard. In this regard, 0.2µl of each selective amplification reaction (approximately 100ng of DNA) proved to be optimal in reactions containing 0.5µl of LIZ1200 size standard.

The inclusion of the LIZ1200 size standard allowed size-determination of fragments up to 1200bp in length. Although resolution decreased slightly at higher fragment sizes, fragments of up to 700bp in length could generally be measured with single base pair resolution. Fragment sizes rarely exceeded

this length when *Mse*I, a restriction enzyme with a 4-bp recognition sequence, was used for DNA digestion.

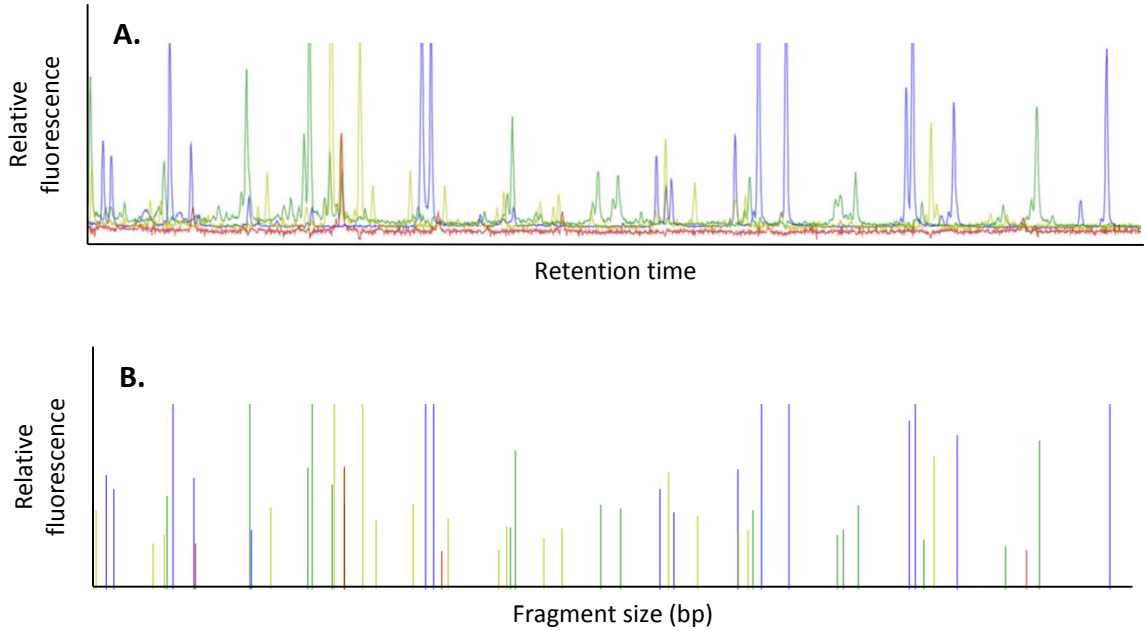


Figure 4.9 A) Raw chromatogram data; B) Peak calls based on a 100 rfu threshold. Peaks were detected by capillary electrophoresis of 4-channel multiplexed S-SAP fragments. Key to colour channels: blue = *Gret1*, green = *Edel*, red = *Cremant*, yellow = *Noble*. LIZ1200 size standard channel (orange) not shown. Region shown is 70-320bp.

Pseudogel imaging software (Genographer V2.1.4), which represents chromatogram peaks as bands in vertical lanes, with band intensity corresponding to peak height, proved an efficient tool for side-by-side comparison of samples (Figure 4.10). Individual lanes can be rearranged and samples from multiple experiments can be directly compared, due to the inclusion of a size standard in each sample.

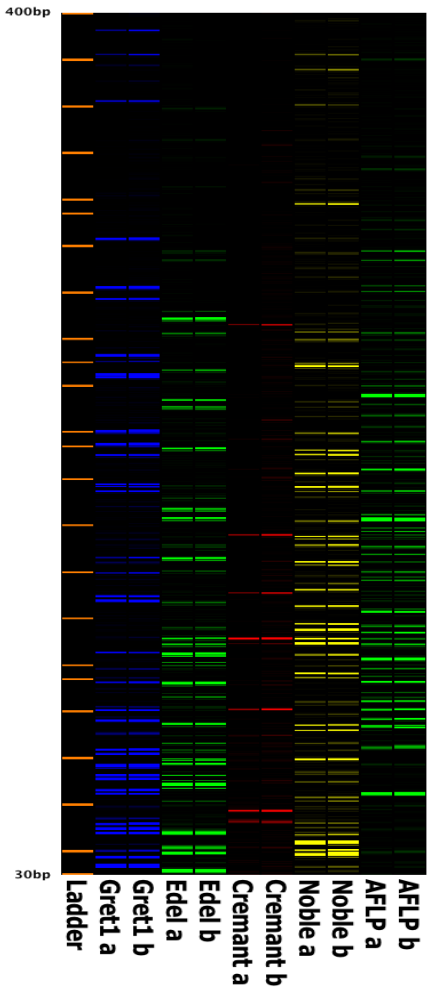


Figure 4.10 S-SAP pseudogel image. Different profiles (lanes) are generated by using retrotransposon-specific primers tagged with different fluorescent dyes. Replicated reactions processed separately show a high degree of profile consistency. Dye colours correspond to the retrotransposon transposon primers used (see lane names). All reactions were performed using Pinot noir UCD13 template DNA and an Mse(GG) adapter primer. Suffix letters (a and b) indicate reactions performed with DNA from two separately processed samples from the same leaf material. Reactions shown in the last two lanes ('AFLP a' and 'AFLP b') show AFLP profiles for comparison; these used fluorescently labelled EcoRI(0) adapter primer in place of retrotransposon-specific primer. Ladder: LIZ1200bp size standard.

The influences of different parameters on S-SAP profiles

As expected, the number of selective nucleotide bases added to the adapter primers had a direct effect on the number of fragments amplified, with additional selective bases decreasing the number of observed fragments. In addition, amplifications performed from the *Eco*RI (6-bp cutter) adapters produced fewer fragments than amplifications from the *Mse*I (4-bp cutter) adapters on co-digested samples (Figure 4.11).

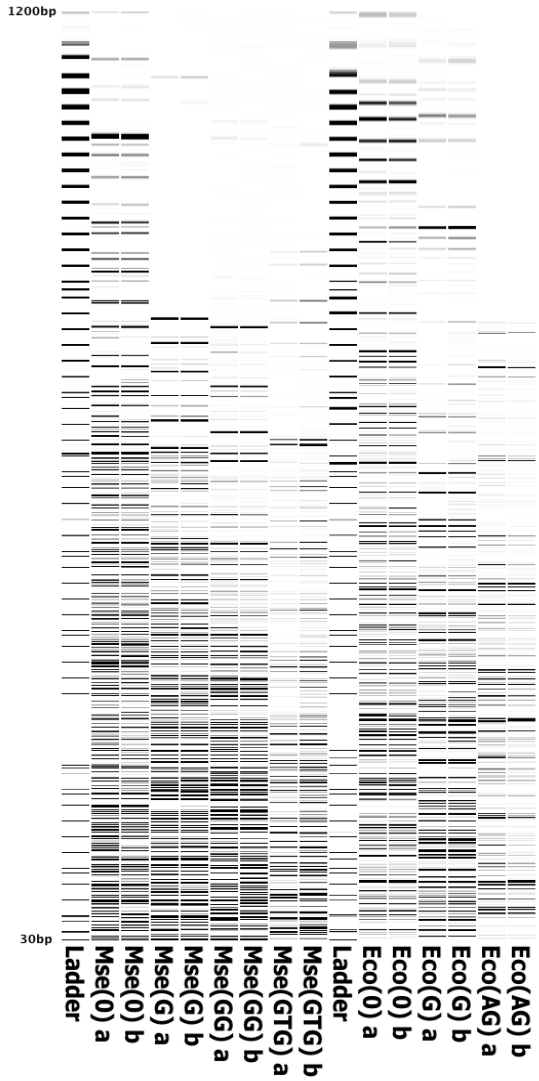


Figure 4.11 The number of selective bases influences S-SAP profile complexity. Amplification profiles were produced using adapter primers with 0, 1, 2 or 3 selective nucleotide bases (indicated in lane names). In each case, VIC-labelled Edel LTR rev(0) primer was used to tag fragments and two separately processed Pinot noir UCD5 DNA samples (indicated by a and b) were used as templates. Ladder: LIZ1200 size standard.

Knox and colleagues (2009) reported improvements to S-SAP throughput by omitting preamplification and adding unlabelled LTR-specific primer to the selective reaction as a third primer. In this study, samples that had not been preamplified did not produce amplification profiles. The inclusion of unlabelled LTR primer and changes in the sample reaction volume also had no effect on the S-SAP profiles (Figure 4.12 A). Altering the $MgCl_2$ concentration in the reaction samples also had no observed effect (Figure 4.12 B).

In contrast, the brand of DNA polymerase used for selective amplification had a strong effect on the resulting fragment profile (Figure 4.12 C). The strongest bands appear in all reactions, although the fragments produced using PrimeSTAR (Takara) are slightly smaller, as this enzyme does not produce fragments that are 3' adenylated. The use of both ExTaq and PrimeSTAR (Takara) enzymes resulted in

fewer bands than Taq (Qiagen). Unlike Taq polymerase both of these enzymes posses 3' to 5' exonuclease activity. The Qiagen Taq polymerase is therefore probably more capable of amplifying fragments to which the primers to not bind with 100% identity. This is important since there is a degree of divergence among the LTRs of different insertions of the same retrotransposon family.

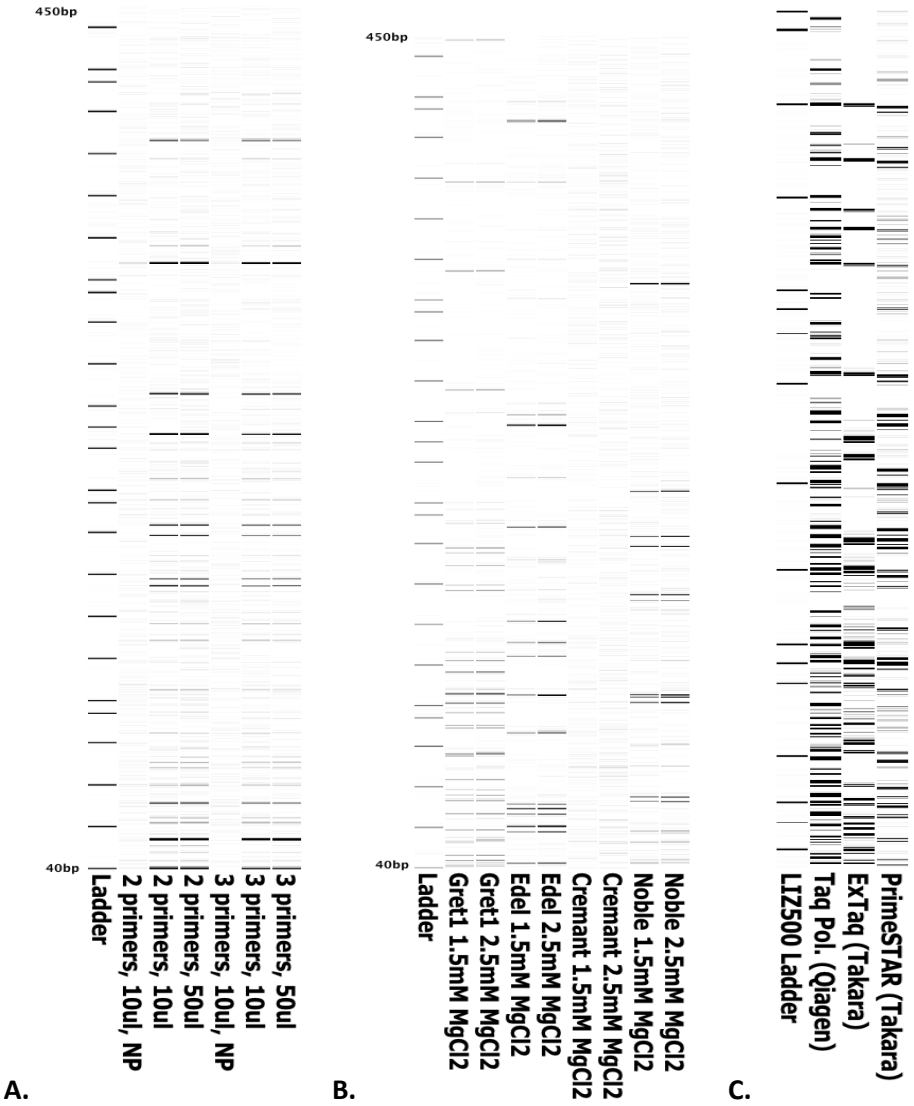


Figure 4.12 The influence of reaction variables on S-SAP profiles. A) The addition of unlabelled LTR primer (3 primers) and changes in total reaction volume (10µl / 50µl) did not affect S-SAP profiles. Performing the selective amplification on non-preamplified (NP) samples was ineffective however. Reactions were performed using Mse(TG) and Cremant LTR rev(0) primers on Pinot noir clone 114 DNA. **B)** The inclusion of $MgCl_2$ at 1.5Mm or 2.5mM did not significantly affect S-SAP profiles. Reactions were performed using Mse(TG) and all four LTR rev(0) primers on Pinot noir clone 114 DNA. **C)** Profile variations resulting from different DNA polymerase enzyme alone. Reactions were performed using Mse(TG) and Gret1 LTR rev(0) primers on Pinot noir clone UCD5 DNA.

4.4.4 Variation between grapevines

Because adapter-specific primers with two selective bases were used for optimised selective amplifications, approximately one sixteenth ($1/4^2$) of the total pool of fragments flanking each retrotransposon family is amplified concurrently during selective amplification. Within this subset of fragments, sufficient profile variation was seen to differentiate each variety based on either *Gret1*, *Edel*, or *Noble* insertion patterns (Figure 4.13 and Appendix B.3). The exception to this was the four Pinot varieties (P. blanc, P. gris, P. noir, P. Meunier). Cremant insertions were too few (less than 10 bands per sample) to allow differentiation of all varieties. While a few clones could be distinguished by unique polymorphisms, others could only be classified into sub-varietal groupings (Figure 4.13).

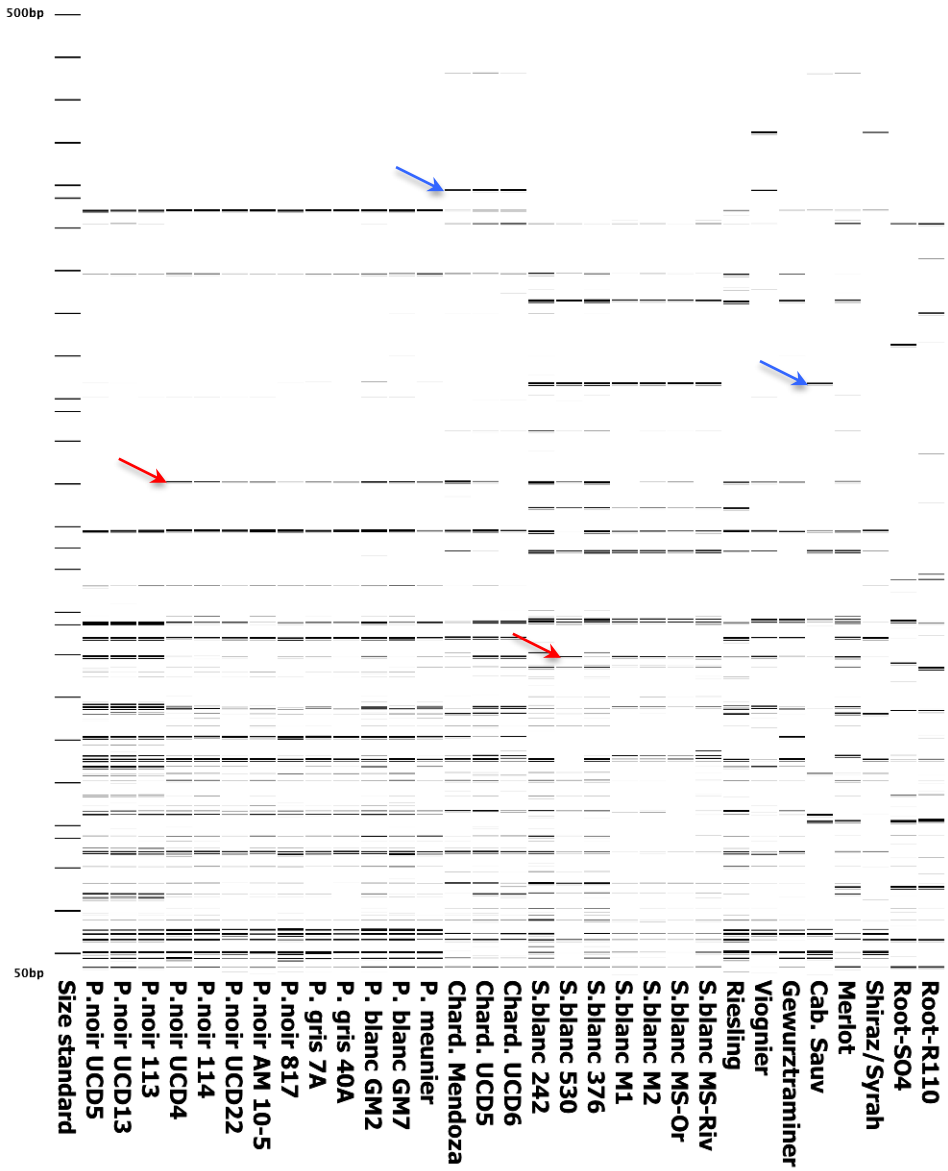


Figure 4.13 Transposon insertion profiles (S-SAP) of different *V. vinifera* clones and varieties. Selective amplification reactions included VIC-labelled Edel LTR rev(0) primer and Mse(TG) primer. Left arrow indicates an insertion characteristic of a subgroup of Pinot noir clones. Red arrows indicate examples of interclonal polymorphisms; blue arrows indicate examples of intervarietal polymorphisms. Profiles generated using primers targeting *Gret1*, *Cremant* and *Noble* retrotransposon families are given in Appendix B.3.

4.4.5 Comparison: fluorescent dye-S-SAP and biotin-S-SAP

Streptavidin bead capture proved effective for selectively purifying fragments tagged with the biotin-labelled primers. Once resolved by PAGE, the fragment profiles clearly matched those produced with fluorescent-labelled primers by capillary electrophoresis (Figure 4.14). It is important to take note of the location of size standards when comparing PAGE with pseudogels, as the separation of fragments by PAGE is not linear with respect to fragment size.

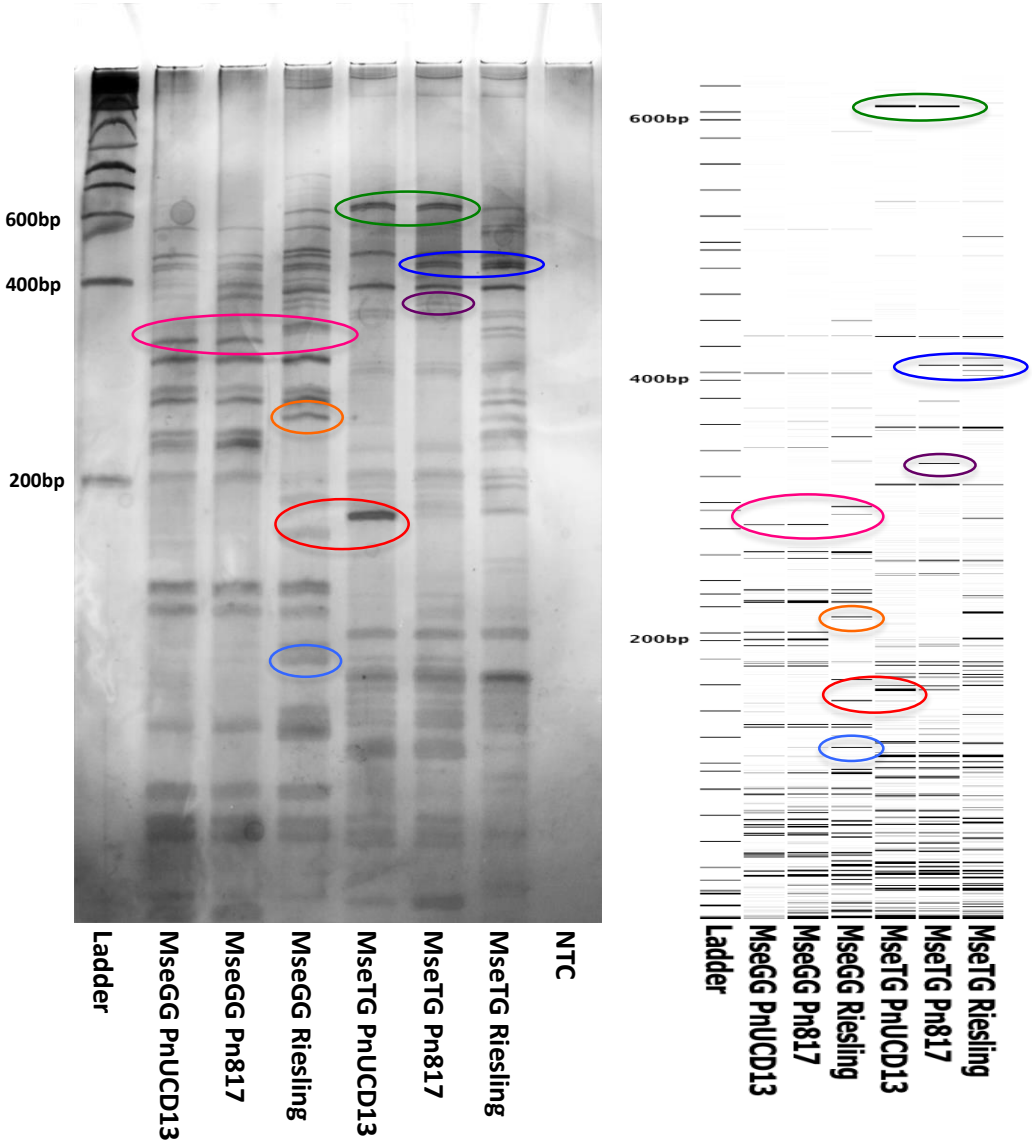


Figure 4.14 Comparison between biotin-capture S-SAP PAGE and fluorescent tagged S-SAP pseudogel. As indicated by the respective ladders, the pseudogel (right) has a linear scale, while the PAGE (left) does not. Nevertheless, it is fairly simple to identify corresponding bands between the two patterns: several are circled in the same colours as an indication. Selective amplifications were performed on the same three samples (Pinot noir UCD13, Pinot noir 817, Riesling) using the Gret1 rev(0) primer and either Mse(GG) or Mse(TG) adapter primer, as indicated on the image. Ladder: Hyperladder I (Bioline) for PAGE, LIZ1200 size standard (ABI) for pseudogel, NTC: no template control.

4.4.6 Individual polymorphic bands can be converted to RBIP markers

Gel punches taken from polyacrylamide gels were effective as templates for PCR amplification in almost all cases (Figure 4.15). Although amplification specificity was checked by agarose gel electrophoresis, contamination resulting from mixed amplification products was occasionally visible in chromatograms produced during sequencing (Figure 4.16), indicating that multiple products may be captured in a single polyacrylamide plug.

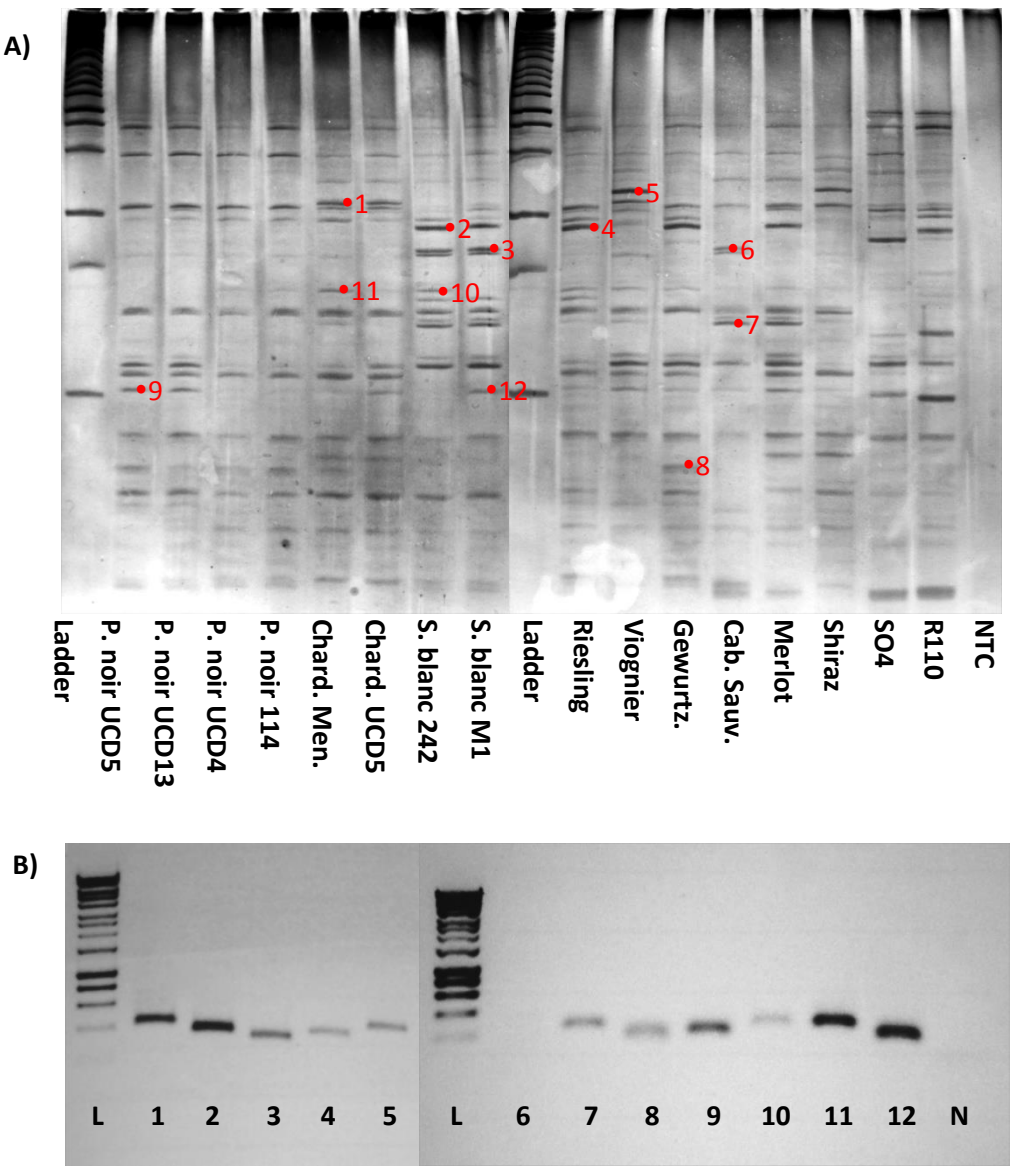


Figure 4.15 Re-amplification of individual S-SAP bands. A) Biotin-tagged S-SAP from sixteen samples amplified using a biotinylated LTR rev(0) primer and Mse(TG) adapter primer. L: Hyperladder I (Bioline), NTC: no template control. **B)** Amplification products from S-SAP gel punch templates. Lane numbers correspond to band identities from A), Ladder: Hyperladder I (Bioline), N: no template control.

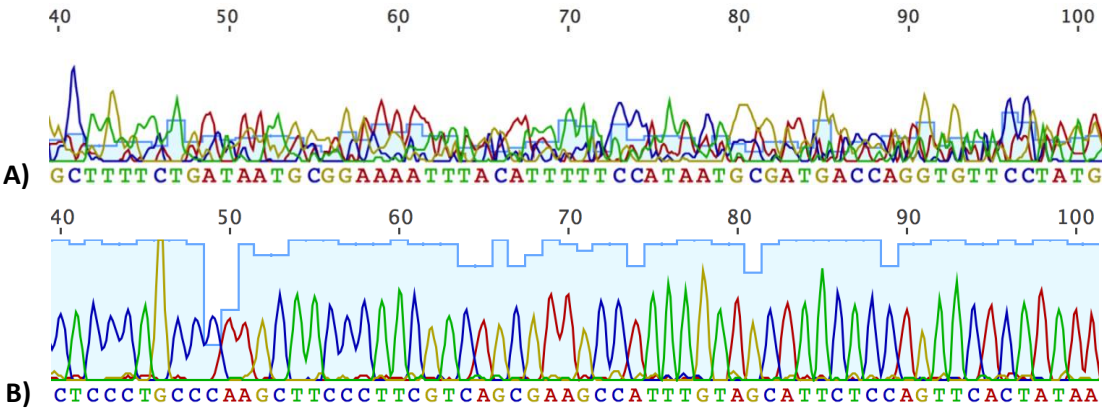


Figure 4.16 Sequence quality checking by chromatogram analysis. Fragments of very similar size may appear as a single band during gel electrophoresis. Although peak quality may be low, sequencing software will often still generate a sequence of base calls. It is essential therefore to check the raw chromatogram data before designing RBIP primers. **A)** Low-quality sequence: 193bp band #7 from Cabernet Sauvignon **B)** High-quality sequence: 299bp band #3 from Sauvignon blanc clone M1. Blue bars indicate the quality scores at each peak (per cent).

Thirty-three polymorphic bands extracted from S-SAP gels produced distinct products when re-amplified and were sequenced by Sanger sequencing. Of these, no sequence data was recovered for 11 of bands that were individually sequenced, due either to contamination of the sequence or the inability of the sequencing software to call distinct peaks (Table 4.2). High-quality band sequence data for the remaining 22 retrotransposon-flanking sequences was mapped to the 12X PN40024 grapevine genome. Of these sequences, 17 could be localised to specific genomic locations. Ten sequences mapped to unique sites in the genome sequence and seven mapped with sufficient specificity to individual copies of repetitive DNA. Of those sequences that could not be mapped, three matched multiple repetitive DNA sequences without significant specificity to any one element and two had no match in the PN40024 grapevine genome database.

The majority (9 out of 11) of the band sequences extracted from S-SAP profiles of P. noir clones mapped to specific retrotransposon insertion sites in the PN40024 reference sequence. For S-SAP reactions involving *Gret1* primers, retrotransposons identified in the grape genome reference sequence adjacent to these loci each proved to be from one of three families: *Gret1*, *Gypsy3*, and *Gypsy16*. The LTR sequence variation across these three families at the *Gret1* primer binding site is minimal (1-2 SNPs in sequence corresponding to the 5' region of the primer). No retrotransposons were present in the reference sequence at loci identified from S-SAP bands isolated from varieties other than P. noir and absent from all P. noir samples.

Table 4.2 Summary of capillary sequence data of recovered transposon-flanking fragments.

Mapping status	Number of fragments	Proportion of fragments
Map to a specific genomic locus	17	65%
• *In unique DNA*	*(10)*	*(38%)*
• *In repetitive DNA*	*(7)*	*(27%)*
Map to multiple locations	3	12%
Do not map to PN40024 genome	2	8%
No sequence data available	11	33%
• *Contaminated sequence (mixed amplicons)*	*(6)*	*(18%)*
• *Sequencing reaction failed*	*(5)*	*(15%)*
Total	33	100%

Primers for RBIP marker analyses were designed on each of the fragment sequences (Figure 4.17). The fragment length in many cases limited optimal primer design, and for one fragment no primer could be designed with sufficiently high annealing temperature. Nevertheless, the RBIP primers obtained proved useful for direct amplification from genomic DNA, such that PCR reactions with each of the 13 primer pairs designed produced single bands of the expected size when tested on samples in which the S-SAP band was observed (Figure 4.18). Reactions performed on DNA from samples in which the associated S-SAP band was absent produced no product (41%), the same product size as observed for positive samples (36%) or a different sized product (23%).

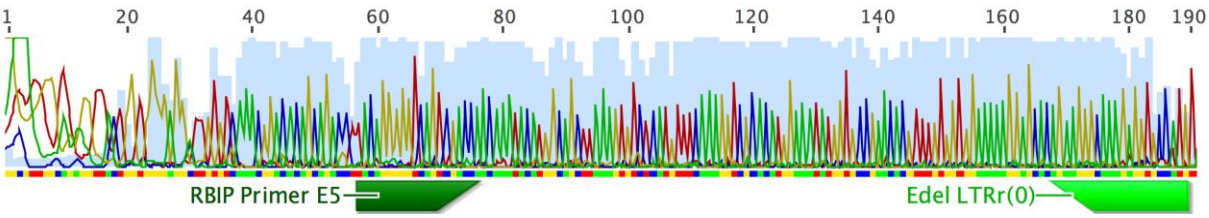


Figure 4.17 Example of RBIP primer locations. Chromatogram sequence data from specific isolated polymorphic bands was used as the template for RBIP primer design. The RBIP primer (dark green) is designed to produce amplification product when used in PCR in combination with the LTR primer which was used to capture the fragment (light green).

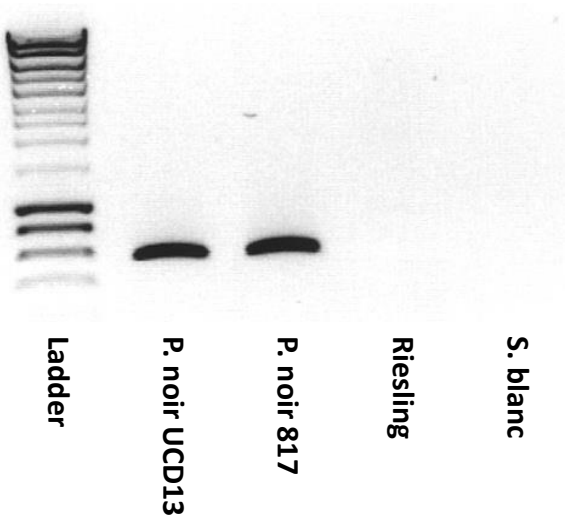


Figure 4.18 Amplification of four genomic DNA samples with a RBIP primer pair. An RBIP primer designed with complementarity to a polymorphic band found only in Pinot noir S-SAP profiles was paired with the Gret1 rev(0) primer used to produce the S-SAP profile.

4.4.7 M-STD reveals the methylation status of DNA adjacent to retrotransposon insertions

Transposon insertion profiles were effectively produced using both *Msp*I and *Hpa*II enzymes for restriction digestion of the DNA. As expected of isoschizomeric enzymes, the banding patterns produced with each enzyme showed a high degree of similarity (Figure 4.19). However, bands appearing in the *Msp*I profiles, but not in the *Hpa*II profiles, indicate methylation of the internal cytosine of the associated CCGG restriction site on both strands. Although polymorphism was seen among the three genotypes tested, only monomorphic bands were scored in order to quantify the conserved epigenetic methylation around TE insertions of various families. No bands present only in *Hpa*II digestions, indicative of hemimethylation of the external cytosine, were conserved across the three clones. The proportion of methyl-sensitivity for sites flanking each of the four retrotransposon families tested in a selective amplification of $1/16^{th}$ of all markers (adapter primer with two selective bases) is given in Table 4.3.

Figure 4.19 Methyl-sensitive transposon display comparison of *V. vinifera* genotypes. Red boxes indicate examples of conserved methylation preventing *Msp*I digestion across all three samples. Three DNA samples (P. noir UCD13, P. noir 818, Riesling) are indicated by lane names. Samples were digested using either *Msp*I or *Hpa*II, followed by S-SAP analysis with fluorescently tagged LTR primers and H/M(CT) adapter primer.

Table 4.3 Methylation sensitivity of monomorphic M-STD markers. Bands appearing in both *Msp*I and *Hpa*II digests are scored as methylation insensitive, while those appearing in only *Msp*I digests are scored as methylation sensitive.

Transposon family	Number of sites		Total
	Methylation insensitive	Methylation sensitive	
Gret1	11 (69%)	5 (31%)	16
Edel	7 (100%)	0 (0%)	7
Cremant	3 (75%)	1 (25%)	4
Noble	15 (88%)	2 (12%)	17
TOTAL	*36 (82%)*	*8 (18%)*	*44*

4.4.8 FTA cards are effective as templates for RBIP screening, but not for S-SAP

Sequence-specific amplification polymorphism profiles generated from DNA preserved on FTA cards showed less than 20 distinct bands per sample following selective amplification. The bands that were observed frequently varied between replicate samples (Figure 4.20). The single-target PCR amplification of RBIP markers produced the same results irrespective of whether purified DNA or prepared FTA card punches were used as a template. The inclusion of a 70% (v/v) ethanol wash step with prior to the manufacturer's recommended wash steps during FTA punch preparation was necessary for efficient PCR from prepared punches.

Figure 4.20 S-SAP amplifications from material stored on FTA cards. S-SAP amplifications performed using material preserved on FTA cards produced few bands that were inconsistent and did not match those produced using purified DNA as a template. Two replicate reactions performed on separate punches off the same FTA card press are shown (FTA1 and FTA2), along with reactions performed on purified DNA for comparison (Purified). MseTG adapter primers were used in each case, in conjunction with the LTR primers specific to the retrotransposon family shown (Gret/Edel; Appendix C). PnUCD5: Pinot noir clone UCD5, SblUCD1: Sauvignon blanc clone UCD1.

4.5 Discussion

4.5.1 The limitations of IRAP and REMAP

Although both IRAP and REMAP involve performing only a single PCR reaction on purified DNA, both techniques have limitations that render them unsuitable as marker systems for use in grapevine. In the case of REMAP very few bands were seen for each primer combination. The appearance of several very darky-staining bands suggests that preferential amplification of certain loci prevents the generation of a large number of features. While IRAP profiles are more complex, the lack of consistency across DNA preparations from the same material makes comparison of individual marker patterns impossible.

Benjak and colleagues (2006) have reported disparities in AFLP profiles that are the result of variations in grapevine DNA extraction techniques. The IRAP and REMAP techniques are comparable to the RAPD technique in that multi-locus amplification is performed without the ligation of adapter oligonucleotides. In this regard it is worth noting that a network of European labs have shown RAPD profiles to have low reproducibility compared to AFLP (JONES *et al.* 1997).

4.5.2 The advantages of S-SAP with fluorescent tags

In contrast to IRAP and REMAP, the S-SAP procedure generated feature-rich banding patterns that were consistently reproducible across multiple DNA extractions from the same sample tissue. The increased reproducibility of this technique may be due to the fact that primers anneal exclusively to synthetic DNA without epigenetic modification or bound proteins. During the preamplification reaction primers bind only to the oligonucleotide adapters. The diluted products of this reaction are then used as templates for selective amplification PCR.

In addition to increased reproducibility, the S-SAP marker system displayed a number of benefits over the other transposon-based genotyping techniques, largely due to the use of fluorescent dye-labelled primers. Since the sizing of labelled fragments is performed by capillary electrophoresis, the marker profiles produced are of high resolution, allowing fragments differing in length by only a single base pair to be discriminated. The high resolution extends across a considerable range of fragment sizes, approximately 30 - 1000bp, which is greater than that which can commonly be resolved by gel electrophoresis.

Furthermore, the results produced are in the format of digital chromatograms, which can be processed, arranged and displayed by computer software. To each sample that is analysed a size standard (comprising a set of fragments of known size, labelled with a unique dye) is added, providing internal markers for size calibration. As a result, profiles generated in separate experiments

can be compared side-by-side. In comparison, gel-based techniques require that samples be run on the same gel in order for fragment sizes to be directly compared.

Finally, up to four fragment populations, produced by the selective amplification of different retrotransposon families, could be analysed simultaneously when labelled with different dye compounds. Such multiplexing directly increases the rate at which marker screening can be performed.

It is important to note that several factors were found to influence the banding patterns generated by the S-SAP technique. The inclusion of a pre-amplification reaction was necessary for the amplification of multiple markers. During the selective amplification reaction, the number of selective bases added to the 3' ends of primers, the concentration of $MgCl_2$ and the brand of Taq polymerase all affected the resulting profiles. Furthermore, the amount of DNA used for capillary electrophoresis had a significant effect on the quality of the chromatogram data.

In contrast, several factors were not seen to alter S-SAP profiles within the ranges tested. For enzymatic digestion of genomic DNA, co-digestion versus sequential digestion by *Eco*RI and *Mse*I restriction enzymes and the incubation period of the reaction (3 hours or overnight) were inconsequential. During selective amplification, the reaction volume (between 10µl and 50µl), Taq polymerase concentration (0.25U-1.0U per 10µl reaction) and the inclusion or omission of unlabelled LTR primer did not affect the S-SAP profiles.

4.5.3 The degree of S-SAP polymorphism between vines

As expected, the similarity between S-SAP profiles of individual vines was reflective of the relatedness of the individuals. Using primers with two selective bases to display one sixteenth of the markers related to a single retrotransposon family, all grape varieties tested could be distinguished by their S-SAP profiles, with the highest variation seen between vines of different species. The exceptions to this observation were the Pinot varieties, which frequently proved to be indistinguishable from one another. Within the tested varieties, inter-clonal variation was also seen. However, clones did not show unique profiles within the subset of retrotransposon elements analysed in a single amplification reaction. Rather, the genotypic variation seen allowed clustering of clone lines into sub-varietal groups.

Since S-SAP bands are produced by the amplification of DNA regions flanking retrotransposon insertions, sequencing polymorphic bands reveals the specific genomic locus affected by the transposon insertion. This information is valuable to the characterisation of genotypic variation that may influence phenotypic differences.

4.5.4 Differences between TE families

For the purpose of genotyping, maximum numbers of polymorphic markers are desired. In complex marker profiles such as those produced by S-SAP, the marker density and the polymorphism rate of the markers determine this. The retrotransposon family targeted by S-SAP primers establishes both of these factors.

Since the activity and age of all retrotransposon families are not equal, those that contribute most to inter-genome variation are most appealing as targets for marker development. In this study, the relative numbers of bands produced for each primer set reflected the number of genomic insertions for those families to which the primers have specificity, indicating that marker density can be estimated from computational searches of the grapevine genome (see 3.4.1). However, the rate of polymorphism among bands must be determined empirically by the comparison of multiple genotypes.

4.5.5 RBIP markers as tools for variety identification

Isolation and re-amplification of individual bands from polyacrylamide S-SAP gels proved to be relatively straightforward. Although the short length of many fragments limited site selection for RBIP primer design, all RBIP primers proved to be effective for direct PCR amplification of genomic DNA in conjunction with the relevant TE-specific primers. However, the polymorphic bands in S-SAP profiles could not be associated with identical RBIP marker amplification in every case. In 36% of the tested cases, PCR amplification of genomic DNA in which the relevant S-SAP band had been absent nevertheless produced a band matching the positive control samples. This could be explained by the presence of a restriction site polymorphism in one sample between the RBIP primer and retrotransposon. This restriction site eliminates a S-SAP band in the related sample, but does not interrupt the amplification of the RBIP product from undigested genomic DNA.

Certain RBIP primers produced spurious products when used to amplify samples that do not produce the associated S-SAP band. This is less problematic, as long as the aberrant products are significantly different in size so as to be distinguished from positive control fragments by gel electrophoresis. Non-target amplification products could also potentially be distinguished from positive control amplification products by HRM analysis.

4.5.6 Epigenetic methylation of retrotransposon-flanking DNA

Transposon display profiles of genomic DNA samples digested with isoschizomers of differing methylation-sensitivity enabled an estimation of the relative methylation status of DNA flanking the four TE classes. These results showed the highest CG methylation of DNA flanking *Gret1* elements

(31%), while none of the seven monomorphic fragments generated using the *Edel* LTR primer showed methylation sensitivity. A recent study by Ocaña and colleagues scored methyl-sensitive markers among Pinot noir clones (2013). Their approach made use of the same restriction enzymes for DNA digestion, but used a second set of AFLP adapter primers rather than transposon-specific primers to anchor the selective amplification. In their study, 21% of monomorphic sites were found to be methylation sensitive, a similar proportion to the methylation sensitive monomorphic sites reported here (18%).

It is important to note that the type of epigenetic modification that can be detected by this approach is limited. Methylation of CCGG sites at both cytosine bases one at least one strand or only the outer cytosine base on both strands blocks digestion by both enzymes used. Fragments associated with these epigenetic contexts will therefore not appear in the data. This is significant as CHG methylation is known to be associated with TE silencing in plants (HSIEH *et al.* 2009).

4.5.7 Increasing throughput: direct genotyping off FTA cards

Tissue collection on Whatman FTA cards is quick and convenient and preserved samples can be stored for long periods at ambient temperature. To prepare samples for testing by PCR, a series of wash steps is used in place of more laborious DNA purification protocols. Consequently, the use of FTA cards for sample collection and preparation greatly reduces the labour and fiscal cost of genotyping.

In this study the reproducibility of simultaneous multiple-target amplifications, as in the case of S-SAP profile generation, was unacceptably low when material preserved on FTA cards was used as a template. Consequently, it is necessary to use purified DNA to identify and isolate individual polymorphisms. However, DNA preserved on FTA cards could be used as a template for amplification of unique target sites. Therefore, once IRBP primers have been designed to target specific polymorphisms, these can be used to amplify DNA preserved on FTA cards to increase the throughput of sample analysis.

4.6 Conclusions

Genetic profiling by S-SAP with fluorescently labelled primers enables reproducible, high-resolution genotyping that is not achievable in grapevine using the IRAP or REMAP protocols. When fluorescent dye tags are replaced with biotin, polymorphic bands can be isolated and sequenced. The sequence data of individual bands can then be used to design RBIP primers, which are suitable for rapid genotyping of leaf tissue stored on Whatman FTA cards. It is important to note, however, that effective RBIP primers cannot be produced for all S-SAP polymorphisms. Bands that appear to be singular in a gel may, in fact, be composed of multiple products, which hinders sequencing without first cloning the amplicons. Furthermore, bands that are polymorphic due to variations at the restriction site cannot be distinguished using RBIP primers designed according to the sequence of the polymorphic band.

In contrast with the microsatellite and SNP variation exploited by earlier marker systems, retrotransposon activity is not regular, with periods of stress capable of increasing rates of transcription and resulting in insertion events (WESSLER 1996; GRANDBASTIEN 1998). As a result, certain retrotransposons display greater polymorphism than other variable loci, and therefore are more informative as genetic markers. Additionally, most retrotransposon families have been found to insert preferentially into euchromatic regions, increasing the likelihood that retrotransposon markers may be associated with phenotypic variation (BRANDES *et al.* 1997; HESLOP-HARRISON *et al.* 1997; PEREIRA *et al.* 2005; TADEGE *et al.* 2008).

The adaptation and modification of existing S-SAP methodologies to produce pseudogel profiles of specific grapevine transposon families has provided a very sensitive and precise tool with which to detect mutagenesis within specific retrotransposon families. Due to the high number of genomic retrotransposon insertions per family, it is necessary to reduce the proportion of markers generated, by adding arbitrary selective bases to the amplification primers. Nevertheless, sufficient polymorphism was seen within the proportion of fragments amplified to allow discrimination of all varieties tested, except among those of the Pinot group. Polymorphic markers were also identified that allowed discrimination of certain clones within the varieties Sauvignon blanc, Chardonnay and Pinot noir. By using methylation-sensitive restriction enzymes at the digestion stage, epigenetic variation around the retrotransposon insertion sites could be compared.

The ability to detect retrotransposon insertion polymorphisms between two grapevine DNA samples amidst a high background of common elements is a crucial step towards the goal of tracking grapevine retrotransposon mobility.

Chapter 5

Improving Transformation Efficiency in Grapevine Tissue

5.1 Overview

To quantify the effect of stress events on transposon activity in grape tissue, it is first necessary to be able to expose cells to stressors in a measured and reproducible way. While the exposure of tissue samples to abiotic stress events such as temperature shock simply requires changing the environmental conditions of the sample, biotic shock exposure requires that organic stressors come into physical contact with living cells, which requires effective penetration of the tissue.

In a comparable process, the direct contact between microorganisms in culture and plant tissue cells is also fundamental to successful bacterial transformation of plants. Transient *Agrobacterium tumefaciens*-mediated transformation of mature plant tissue has become a standard technique for investigation of gene expression and function *in vivo*. In grapevine, the efficacy of transient transformation is limited by the penetration of bacterial suspension cultures into tissue. Researchers have reported the use of vacuum infiltration apparatus or site-specific syringe infiltration to improve penetration of bacterial cultures. However, difficulties in isolating specific portions of a living plant make the application of a vacuum unsuitable for most transient infiltration experiments. Additionally, syringe infiltration produces uneven tissue penetration and frequently results in wounding at the infiltration site, possibly confounding gene expression results.

In this chapter the development of an approach to improve the efficiency of transient transformation of grapevine tissue is descried, improving the toolset available to molecular biologists working with this species. Furthermore, this protocol provided a quantifiable method to optimise the consistent exposure of grapevine tissue cells to microorganisms without wounding, a necessity for experiments involving the biotic stressing of grapevine tissues described later (see Chapter 6).

The transformation efficiency improvements are a result of the addition of low concentrations of a commercially available organosilicone surfactant hereafter referred to by its trade name, Pulse (Nufarm Australia Ltd.). When added to aqueous solutions at low concentrations (0.01% - 0.10% v/v), Pulse greatly increases flooding of intercellular air spaces of submerged plant tissue. The addition of Pulse to *A. tumefaciens* suspensions was tested as an alternative approach to achieve bacterial infiltration at ambient pressure.

To compare transformation efficiencies, the transcription factor VvMYBA1, under the control of an ectopic promoter, was used as a reporter transgene. Expression of VvMYBA1 indirectly promotes the

biosynthesis of anthocyanin, producing red pigmentation in grapevine tissue (see 2.2.10). The use of VvMYBA1 as a reporter allowed the non-destructive quantification of transformed cells in target tissue. The measurement of anthocyanin accumulation also proved to be a more precise measurement of transformation efficiencies, when compared with the conventional ß-glucuronidase (GUS) gene reporter system, and did not require samples to be processed before scoring.

Addition of Pulse was found to have an optimal increase on the efficiency of transformation of young Sauvignon blanc (72-fold) when tissue was dipped for 10 minutes in *Agrobacterium* suspensions with Pulse added at 0.03% (v/v) concentration. Transformation efficiencies of leaves from eight other commercial wine grape varieties and one rootstock variety were also significantly increased under these conditions. Tween 20 and Triton X-100 did not prove to be suitable alternatives. The use of Pulse as an additive is likely to improve *Agrobacterium*-mediated transformation of other plant tissues where penetration is a limiting factor.

5.2 Introduction

5.2.1 Agrobacterium plant transformation

The ability of *Agrobacterium tumefaciens* to transfer portions of its own DNA into the nuclei of somatic plant cells has been widely utilized as a tool for the generation of transgenic plants for research and the biotech industry. Although generation of stable transgenic lines can take months to achieve, transient transformation can be accomplished by direct infiltration of mature tissue with *Agrobacterium* culture (JANSSEN AND GARDNER 1990). Such transiently expressed foreign DNA has proved useful for the *in vivo* characterisation of foreign genes, promoter elements and silencing constructs (GELVIN 2005).

Although grapevine is ranked as the world's most economically important fruit crop, transient transformation of leaf tissue was not successfully reported until 2008 (SANTOS-ROSA *et al.* 2008; ZOTTINI *et al.* 2008). Both of these early reports highlight great variation in transformation efficiencies between leaves from different *V. vinifera* varieties.

5.2.2 Surface tension: the major barrier to bacterial infiltration

The first obstacle to successful transformation via infiltration is the effective penetration of bacterial suspensions into the air spaces of the mesophyll layer of the leaf. Penetration via the stomata is limited by the surface tension of aqueous solutions (approx. 76mN/m). Both of the aforementioned techniques report the application of external pressure, either through the rapid release of a vacuum or by direct syringe infiltration, to introduce the bacterial suspension. In the current study, we show that the addition of an organosilicone surfactant solution containing 1,020g/L organomodified polydimethyl siloxane, marketed as Pulse® Penetrant (Nufarm Australia Ltd.), to bacterial suspensions allows flooding of the intercellular air spaces of young grape leaf samples without the application of external pressure. Such solutions have vey low surface tensions (22mN/m, lower than that of ethanol or acetone) which permits them to flow though stomata, thereby gaining entry into plant tissues (ZABKIEWICZ 2008).

5.2.3 GUS / VvMYBA1 as reporters

In addition to the ß-glucuronidase (GUS) gene (JEFFERSON *et al.* 1987), we used the grape transcription factor VvMYBA1 that regulates the final stages of anthocyanin biosynthesis (KOBAYASHI *et al.* 2005) as a reporter of successful transformation.. VvMYBA1 had previously been used as a reported for promoter activity in both grapevine somatic embryos and tobacco (LI *et al.* 2011). In this study, a construct was used that comprises a VvMYBA1 gene under control of a *cauliflower mosaic virus* 35S promoter (CaMV 35S). Constitutive expression of the VvMYBA1 gene in grape leaf tissue resulted in an accumulation of anthocyanin. The red pigmentation of transformed cells provides high contrast

against the green untransformed tissue, allowing transformation efficiency to be quantified by non-destructive visualisation.

5.3 Methods

5.3.1 Leaf tissue penetration

A dilution series of Pulse in water (0%, 0.01%, 0.03%, 0.10%, 0.50%, 1.00% v/v) was prepared in 50ml Falcon tubes. Young leaves (approx. 5cm wide) of potted Pinot noir vines were submerged in the solution for 10 minutes while on the vine. Leaves were then immediately blotted dry with paper towel, removed and photographed against a black background.

5.3.2 Vector construction

A GUS/GFP fusion gene was inserted into the *SmaI*/*BamHI* sites of the plasmid pART7 directly downstream of the CaMV 35S promoter to form pART7:GUS/GFP. The *NotI* fragment from pART7:GUS/GFP, containing the 35S promoter and GUS/GFP reporter gene, was then inserted into the unique *NotI* site of the binary vector pART27 (GLEAVE 1992) to form the expression vector pART27:GUS/GFP (Appendix A.1.1).

Likewise, the Gateway cassette RFA (Gateway® Vector Conversion System; Invitrogen), which contains the chloramphenicol resistance gene and ccdB gene, flanked by Gateway *attR* recombination sites, was cloned into the *SmaI*/*BamHI* sites pART7 directly downstream of the CaMV 35S promoter to form pART7:GW. The *NotI* fragment from pART7:GW, consisting of the 35S promoter and Gateway casettewas then cloned into the *NotI* site of into pART27 create the plasmid pART27:GW.

The grapevine transcription factor gene VvMYBA1 was amplified from Pinot noir post-veraison berry skin cDNA using forward primer VvMYBA1 GWF and reverse primer VvMYBA1 R (Appendix C) After gel purification, the product was directionally cloned into the pENTR/D-TOPO entry vector (Invitrogen) according the manufacturer's recommended protocol to create pENTR:VvMybA1. An LR Clonase II (Invitrogen) recombination reaction performed between pENTR:VvMybA1 vector and pART27:GW, according to the manufacturer's instructions, produced the expression vector pART27:VvMYBA1 (Appendix A.2). Both expression vectors were transformed into *Agrobacterium tumefaciens* strain GV3101 cells by electroporation (see Appendix E.3).

5.3.3 Transient transformation at ambient temperature

A single colony pick of *A. tumefaciens* containing the vector pART27:VvMYBA1 was grown for 48hr at 28°C, 200RPM in 50ml LB broth with 50mg/l spectinomycin. Bacterial cells were pelleted by centrifugation and resuspended in 100ml liquid medium containing 1X MS vitamins & salts (MURASHIGE AND SKOOG 1962), 30g/L sucrose and 150mM Acetosyringone. After 2hr incubation (28°C, 200RPM), OD_{600nm} was adjusted to 0.8 with additional liquid medium. A dilution series of Pulse (0%,

0.01%, 0.03%, 0.10% v/v) in 30ml of *Agrobacterium* suspension was prepared in four 50ml Falcon tubes. Twenty sterile leaf discs (approximately 1cm^2) from *in vitro*-grown Sauvignon blanc plantlets were added to each tube and tubes were inverted five times to submerge samples. Five randomly chosen disks were removed from each tube at 5-minute intervals. Disks were blotted dry on sterile filter paper and incubated (25°C, 16hr:8hr light/dark cycle) on nutrient plates containing 1X MS vitamins & salt, 30g/L sucrose and 7g/L phyto agar (Duchefa, Haarlem, The Netherlands). After 2 days, leaf disks were moved to plates containing the same nutrient medium, supplemented with cefotaxime (200mg/L). Parallel tests were carried out using two other common laboratory surfactants (Tween 20 and Triton X-100) in place of Pulse for comparison.

To compare transformation efficacy across multiple varieties, young sterile leaves of eight *V. vinifera* varieties and the rootstock R110 (*V. berlandieri* x *V. rupestris*) were infiltrated with *A. tumefaciens* cells carrying the binary vector pART27:GUS/GFP using the above protocol (0.03% v/v Pulse concentration, 10 minute dip duration, n=4). Staining and assays of GUS expression were performed as described in Appendix E.6.

5.3.4 VvMYBA1 efficiency scoring

One week after *Agrobacterium* infiltration, leaf disks were photographed at 12.5-megapixel resolution using a DP71 digital camera mounted on a SZX16 stereomicroscope (Olympus). For each leaf disk, a central region of the disk that showed typical expression pattern and consistent focal distance was photographed at 30X magnification. Five zones on different replicate disks were photographed for each condition. Image analysis software (GIMP v2.6, www.gimp.org) was used to score the number red pixels in each image. Any portions of the image containing cells which showed browning were first masked, and then red pixels were selected using the 'select by colour' tool. Transformation efficiency was calculated by dividing the number of red pixels by the total number of pixels (12,533,760) for each image to produce a percentage score. Two-way factorial ANOVA (time, concentration) was performed on the raw data (n = 5), followed by post-hoc Tukey's test at 95% confidence (TUKEY 1949).

5.4 Results

5.4.1 The addition of Pulse increases *in vivo* infiltration at atmospheric pressure.

An increase in Pulse concentration (up to 1.0% v/v) proportionately increased unassisted flooding of intercellular air spaces of leaves from nursery-grown Pinot noir vines with aqueous solutions, as observed by increasing tissue translucence (Figure 5.1).

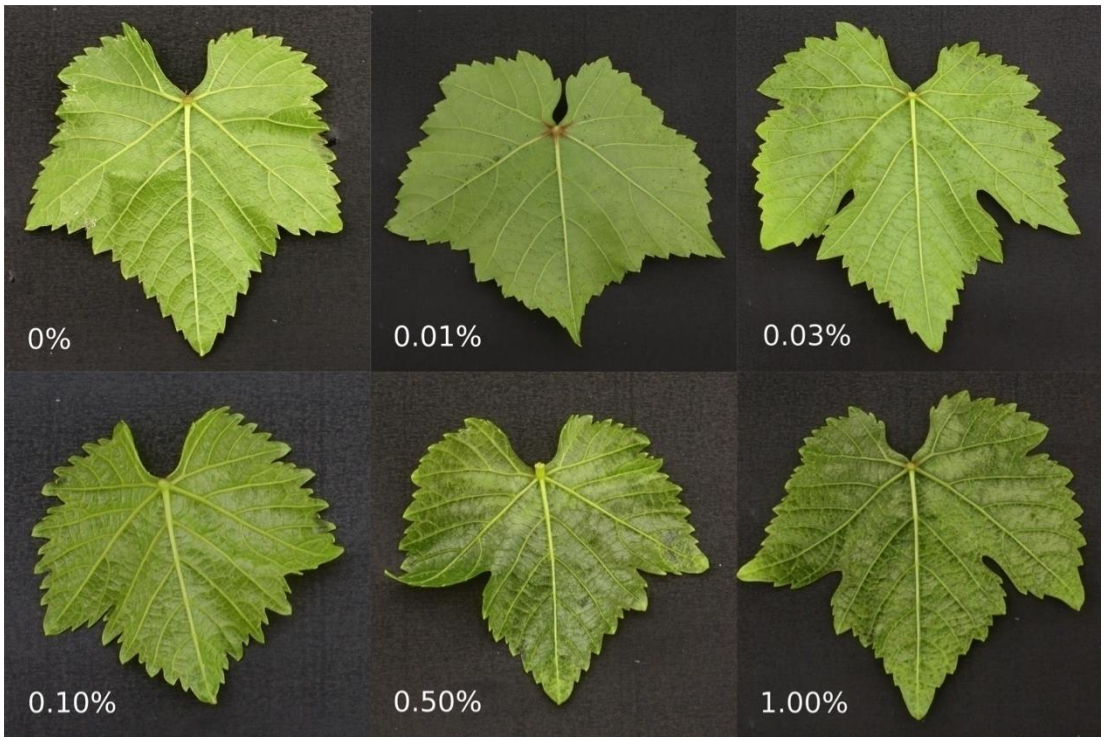


Figure 5.1 Penetration of organosilicone surfactant solution into leaves of nursery-grown Pinot noir vines. After 10 minutes flooding of intercellular air spaces is visible as an increase in tissue translucency (percentages indicate Pulse concentration).

5.4.2 VvMYBA1 is a more precise transformation marker than GUS in grapevine leaves.

Microscopy revealed that in grapevine leaf tissue infiltrated with *A. tumefaciens* carrying the plasmid pART27:VvMYBA1 a proportion of the cells accumulated dark red pigmentation after one week. Transformed cells could be easily discriminated from green wild type cells by their high colour contrast. The diffusion of marker pigment across cellular membranes that is generally seen after GUS staining was not apparent for tissue accumulating anthocyanin (Figure 5.2).

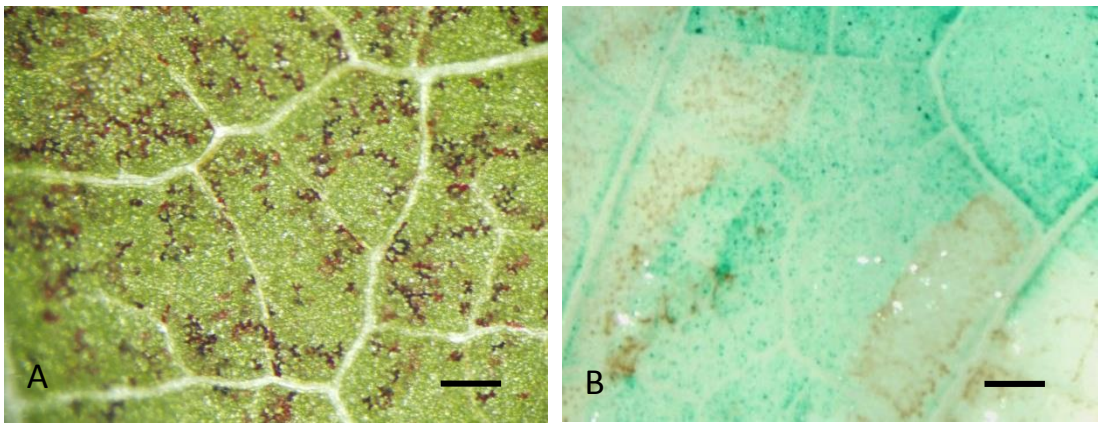


Figure 5.2 Transient transformation of Sauvignon blanc leaves with different transgenes. Each transformation was performed using a 10 minute *A. tumefaciens* dip, with 0.01% (v/v) added Pulse. **A**: Transformation with pART27:VvMYBA1; **B**: Transformation with pART27:GUS/GFP following GUS staining and ethanol destaining. Scale bars represent 0.2mm.

5.4.3 Improved transient transformation

For tissue transformation with pART27:VvMYBA1, transformation efficiency was scored as the proportion of red pixels within a single 12-megapixel frame of leaf tissue (see 5.3.4). The addition of Pulse surfactant to bacterial suspension dips resulted in an average increase in transformation efficiency of Sauvignon blanc leaf tissue from 0.02% (SD=0.02%) after 20 minutes to 1.44% (SD = 0.39%) after 10 minutes (Figure 5.3). Solutions including Pulse at 1.0% (v/v) concentration and prolonged dip duration resulted in browning of some cells (Appendix B.4). As a result, the proportion of cells that could be scored as successful transformations was reduced.

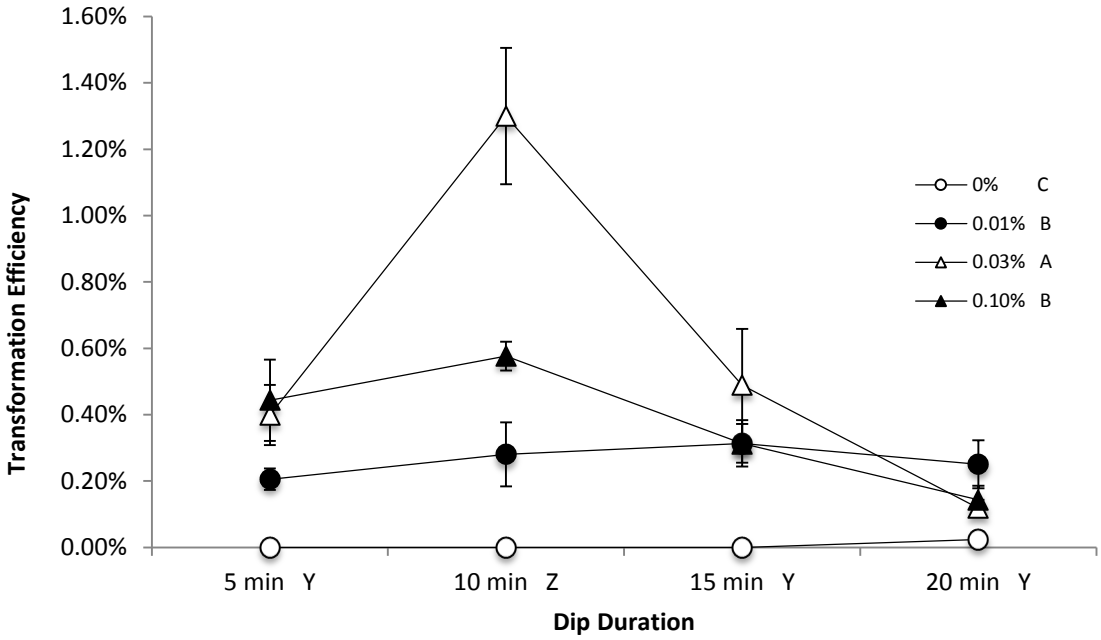


Figure 5.3 The effect of time and Pulse concentration on transformation efficiencies.

Transformation efficiencies of in vitro grown grapevine leaves were found to be optimal when Pulse was added at a concentration of 0.03% (v/v) and dip times were 10 minutes. Error bars indicate standard error of the mean, n = 5. Concentrations (legend) followed by the same letter (A, B, C) and durations (horizontal axis) followed by the same letter (Y, Z) are not significantly different across all samples (Tukey test at 95.0% confidence).

Both dip time and surfactant concentration was found to have a significant effect on transformation efficiencies ($p < 0.001$; Figure 5.3). The 10-minute, 0.03% (v/v) concentration condition that gave the highest average transformation efficiencies was found to be statistically different from all other conditions using Tukey's method at 95% confidence.

Tween 20 did not have an effect on transformation efficiency at concentrations up to 0.10%. Triton X-100 increased transformation efficiencies above those of control samples only when added at a concentration of 0.10% or more, but produced tissue necrosis after one week (Figure 5.4).

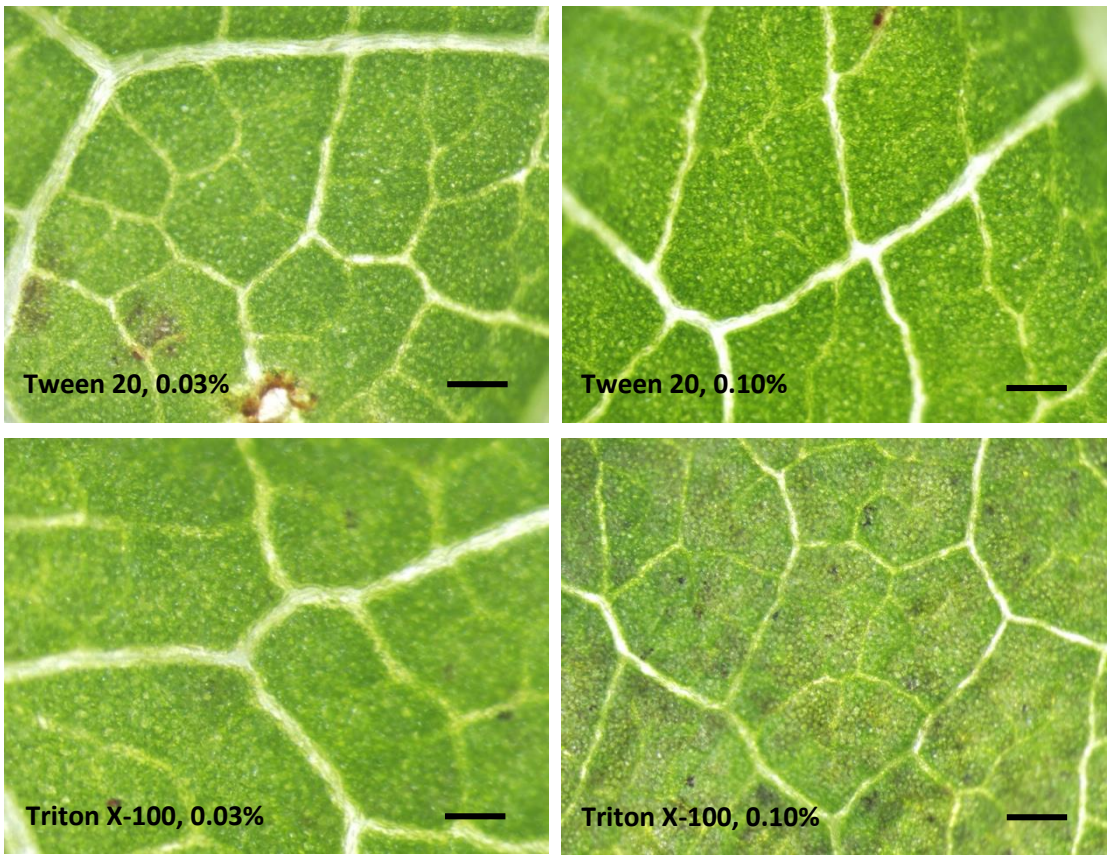


Figure 5.4 Other common laboratory surfactants were not found to be suitable for improving *Agrobacterium*-mediated transformation efficiency. Tween 20 was not effective in raising transformation efficiencies. Triton X-100 was able to slightly increase transformation efficiency at 0.10% concentration, but resulted in browning of leaf tissue. Surfactants (concentrations indicated on each image) were used to supplement *A. tumefaciens* suspensions during dip transformation of S. blanc leaf disks with pART27:VvMybA1. Scale bars represent 0.2mm.

Although transformation efficiencies could not be scored with high precision or accuracy following transformation with pART27:GUS/GFP, the blue-stained tissue is clearly visible without magnification, allowing rapid comparison of samples. Visual analysis of leaf tissue from nine different grapevine varieties transformed by *Agrobacterium* infiltration with the pART27:GUS/GFP expression vector clearly shows that supplementing bacterial suspensions with Pulse elevated transformation efficiencies in each variety tested (Figure 5.5).

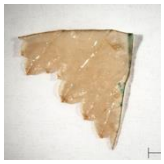
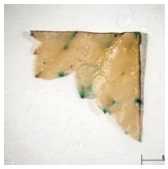
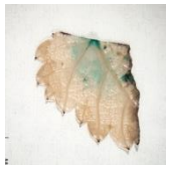
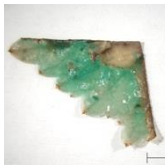
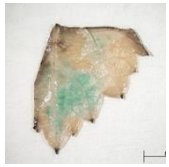
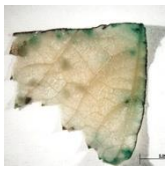
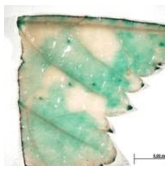
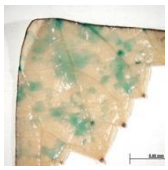
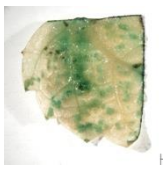
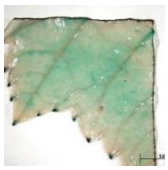
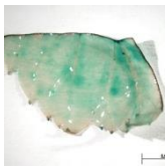
Cultivar:	S. blanc	P. blanc	P. gris	P. noir	Cab. Sauv.
- Pulse					
+ Pulse (0.01%)					
Cultivar:	Chardonnay	Merlot	Shiraz	R110	
- Pulse					
+ Pulse (0.01%)					

Figure 5.5 Transient transformation of nine *Vitis* varieties is improved by the addition of Pulse surfactant. Young leaf tissue samples were transiently transformed with pART27:GUS/GFP via the *A. tumefaciens* dip protocol (- Pulse: without Pulse; + Pulse: supplemented with Pulse). After 48 hours, leaves were stained with GUS stain, destained in ethanol and photographed. Each image is representative of four biological replicates.

5.5 Discussion

5.5.1 The necessity for improved transient infiltration efficiencies

Transient tissue transformation allows the activity of foreign and artificial genes and transcription regulators to be studied within the complex environment of living cells in a matter of days. For this reason, it has become both a popular and invaluable tool for functional genomics research.

In the context of this project, the ability to perform transient transformation of grapevine leaf tissue with good efficiency is important for two reasons. First, the fusion of retrotransposon LTRs with reported genes will be used to test the efficacy of these elements for promoting transcription in somatic tissues of grapevine and tobacco (Chapter 6). Second, the interaction between grapevine tissue and *Agrobacterium* during transient transformation is analogous to the integration between grapevine tissue and biotic stressors (used for TE stimulation in Chapter 6). Just as the efficiency of transient transformation is limited by the number of plant cells with which plasmid-carrying bacteria come into contact, so too is the magnitude of stimuli induced by other microorganisms likely to be related to the exposure of tissue from the two species.

5.5.2 VvMYBA1 gene expression is suitable for quantifying transformation efficiency

For the purpose of measuring transformation efficiencies, the ectopic expression of the grapevine transcription factor VvMYBA1 proved to be highly useful as a reporter gene. The anthocyanin produced in response to its expression accumulates in individual cells and does not diffuse across cellular membranes as is seen in GUS assays. This serves to improve quantification precision. In addition, samples could be assayed directly, without a need for destructive staining steps or any other processing of the tissue.

Colour histogram analysis of leaf tissue using the mean red brightness (MRB) method (Li *et al.* 2011) did not prove reliable, with regions that showed no red cells inconsistently showing ether higher or lower MRB scores than regions which contained red cells (data not shown). This was found to be because brightness increases the MRB score in the red colour channel irrespective of the base colour, such that white has a maximum MRB score of 255 and black has a score of 0. Light green coloured tissue therefore has a higher MRB score than dark green tissue in the same image. In contrast, the select by colour tool takes into account the green and blue channel scores when selecting pixels. Selecting red pixels in microscope images proved to be simple and reliable, due to the high colour contrast with the green tissue not expressing anthocyanin.

5.5.3 Improved transient transformation

At ambient pressure, supplementing bacterial suspensions with a low concentration (0.03% v/v) of surfactant increased peak transformation efficiencies by an average of 72-fold when tissue was submerged for only 10 minutes. Further increases to Pulse concentration or exposure time reduced transformation efficiencies, as a result of tissue damage.

As others have reported, the efficiency of grapevine leaf transformation differs greatly between cultivars (SANTOS-ROSA *et al.* 2008). However, the addition of Pulse improved efficiency in all cultivars tested. Most notable were the efficacies obtained for S. blanc and P. noir, two economically important varieties which showed almost no transformation without Pulse. When considering the variation in efficiencies observed between leaf tissue samples from multiple *Vitis* varieties, it is worth noting that inter-varietal variation in leaf structure is well known, and is in fact a trait commonly used for variety identification.

Vacuum infiltration, currently the most common transient transformation procedure, necessitates the use of a pump and vacuum chamber and is impractical when explant tissue cannot be isolated, such as in the case of leaves attached to the vine. The current alternative technique, syringe infiltration, results in uneven saturation of the tissue and often causes wounding at the infiltration site. A simple bacterial suspension dip protocol, supplemented with Pulse to increase tissue penetration, is less technically challenging than previously published grapevine transient transformation methods and avoids the possible confounding effects of tissue wounding.

5.6 Conclusions

The addition of an organosilicone surfactant to *Agrobacterium* suspension cultures increases the flooding of intercellular air spaces of grapevine leaves at ambient pressures. Transient transformation efficiencies were improved 72-fold without the need for vacuum or syringe infiltration apparatus, which can be impractical or cause wounding. It is likely that the addition of Pulse would improve transformation efficiencies in other plant systems where penetration of culture into the tissue is a limiting factor. Due to the apparent intervarietal variation, it is recommended that dip times and efficiencies first be optimised to each cultivar.

The VvMYBA1 gene was found to be useful as a reporter of transformation efficiency, due to the high colour contrast between cells that produced anthocyanin and those that did not. The lack of anthocyanin trafficking between cells and the non-destructive nature of the assay allowed easier and more precise quantification of transformation efficiency in tissue transformed with VvMYBA1 versus the GUS reporter gene.

Chapter 6

Transcriptional Regulation of Grape Retrotransposon Activity in Response to Environmental Stress Events

6.1 Overview

The abundance and diversity of retrotransposon families in the modern grape genome (Chapter 3) provides a resource for the development of molecular markers by which varieties and clones can be differentiated (Chapter 4). This legacy of retrotransposon-based variation indicates that much of the genetic variation within this species is driven by the activity of these elements. In contrast to sexual recombination, which is limited to a specific event, somatic mutations can potentially accumulate at any stage of the plant's life cycle, in any tissue. Furthermore, external forces may affect the rate at which these mutations accumulate. Several plant retrotransposons have been shown to have increased activity following environmental stress events (see 2.2.5) but little is known about the effect of changing environmental conditions on transposon mobility in grapevine.

Transcription is a necessary first step for transposition, and is often the stage at which a host genome supresses transposon activity. In the experiments described in this chapter, quantitative PCR was used to measure the increased transcription of four retrotransposon families (*Gret1*, *Edel*, *Cremant* and *Noble*) following particular stress events. Elevated transcription of transposable elements (TEs) was observed in stressed grapevine embryogenic callus (EC) cultures, but not in the leaf tissue of plantlets in tissue culture or potted, nursery-grown vines. Grapevine EC cultures exposed to cultures of yeast isolates collected from local vineyards showed greatly increased retrotransposon expression, the like of which was not seen in cultures exposed to abiotic stresses. An observed increase across multiple retrotransposon families suggests a change to the silencing of endogenous retrotransposons in this tissue following yeast stress.

Constructs comprising a GUS reporter gene under transcriptional regulation of retrotransposon LTRs showed the ability of three out of four of these sequences to induce transcription within a model organism, *Nicotiana tabaccum*. Stable transgenic *N. tabaccum* plants did not show increased LTR-reporter transcription following infiltration with yeast cultures, although wounding appeared to moderately increase transcription in the case of *Cremant* and *Noble*. Introduction of the LTR-reporter constructs into *V. vinifera* leaves showed that three out of four of the retrotransposon LTRs tested are not able to induce expression in this tissue. Expression was only seen from the LTR of *Cremant*, the least abundant of the four retrotransposons in the grapevine genome.

These experiments demonstrate the host-specific repression of transcription for the four retrotransposon families studied. The elements are, however, able to become transcriptionally active following specific biotic stresses, but this activation was specific to embryogenic calli growing on solid medium.

6.2 Introduction

The grapevine genome contains an abundance of functionally complete retrotransposons (see 3.5.1). Although S-SAP profiles show that retrotransposition has resulted in genetic variation between varieties, the low degree of polymorphism between clones indicates that the mobilisation of these elements, and hence their mutagenic capacity, is limited under normal conditions. This is believed to be true of TEs in general, whose activity would otherwise be genotoxic to the host organism.

As described, retrotransposition requires the transcription of existing elements followed by the integration of reverse-transcribed copies into the genome (see 2.2.2). Transcription of retrotransposons is therefore necessary but not sufficient for the generation of new insertions. Consequently, the conditional and tissue-specific expression patterns of retrotransposons within an organism place spatial and temporal limitations on potential retrotransposition.

Measuring variations in the non-constitutive expression of grapevine retrotransposon families provides information regarding the tissue types and circumstances under which retrotransposon transcript abundance is increased. For example, the transcription of the tobacco retrotransposon Tnt1 has been shown to correlate with the accumulation of new insertions (MELAYAH *et al.* 2001). This activity can then be used to determine conditions under which retrotransposon-associated mutagenesis may occur and those families that are most likely to be associated with new mutations.

6.2.1 Grapevine TE responses to stresses

Stress response *cis*-regulatory elements (CREs) are numerous in the retrotransposon sequences of grapevine (see 3.4.3), and other plants (reviewed in 2.2.5). These motifs provide binding sites for transcription factor proteins, which help initiate mRNA production. The association between stress-response CREs found in retrotransposons and stress response genes suggests that retrotransposon transcription may be increased when grapevine tissues are exposed to stress events.

In addition to transcriptional regulation by transcription factor proteins, TE activity is also controlled by DNA methylation (see 2.2.4). Baranek and colleagues (2010) used an AFLP-based technique called methylation-sensitive sequence amplified polymorphism (M-SAP) to demonstrate that *in vitro* cultivation and thermotherapy treatments result in epigenetic changes to the DNA of grapevine somatic tissue. Their experiments found that the methylation status of approximately one-fifth of all CCGG loci scored was altered by the treatments, with a prevalence for decreased methylation. A general decrease in DNA methylation of grape tissues in response to stress events may, at least in part, contribute to an overall increase in transposon activity.

6.2.2 Tissue types

Expression of transposons in specific tissue types has also been reported (POUTEAU *et al.* 1991; HU *et al.* 1995; TURCICH *et al.* 1996). Sequence analysis of five grapevine retrotransposon families shows regulatory motifs associated with embryogenesis, seed-specific expression and meristematic tissue-specific expression (Table 3.3). Since the goals of this project involve regeneration of whole plants (see 1.6.5), retrotransposon transcript levels were measured in totipotent EC cells under normal growth conditions and after exposure to a range of stress treatments. Cell cultures maintained on both solid nutrient plates and as suspension liquid cultures were tested. For comparison, the retrotransposon expression response of leaf tissue from nursery-grown vines (attached leaves) and *in vitro*-grown vine plantlets (detached leaves) was also tested before and after stress exposure.

6.2.3 Stress types

Grapevine EC were exposed to several abiotic stresses that have been shown to induce retrotransposon activity. While these treatments were outside of ideal growth conditions, the aim was not to apply selection for resistance to the tissue. For this reason, the shock conditions selected were well within the tolerance ranges of the tissue. In each case, the tissues were harvested 48hrs after initiation of the shock exposure. For high and low temperature shock conditions, cells were maintained at 37°C and 4°C respectively for two days. Cells were also treated with short wave radiation by briefly exposing the cells to UV-C and then allowing the cells to recover for two days under normal growth conditions.

To simulate biotic stresses to which *V. vinifera* may be exposed, microorganisms that live in symbiotic relationship with grapevines were used. A strain of the common pathogenic fungus, *Botrytis cinerea*, and six wild yeast species (*Saccharomyces cerevisiae, Pichia kluveri, Aureobasidium pullulans, Hanseniapora uvarum, Rhodotorula glutinis, Cryptococcus magnus*) were collected from New Zealand vineyards. These yeast species (collected by Matt Goddard, Auckland University) represent six of the most common endemic yeasts found in New Zealand vineyards. Two other biotic stressors were also chosen. The first, Onozuka R-10, is a crude cellulase extract from *Trichoderma viride* commonly used in for protoplast isolation that has been found to induce the activity of tobacco retrotransposons (see 2.2.5). The second, the bacterium *A. tumefaciens*, is a known pathogen of Eudicots, responsible for Crown Gall disease. It is also commonly used as a tool for the artificial genetic transformation of plants (including grapevine). This bacterium was included as an example of a prokaryotic pathogen, and to determine whether certain transgenesis protocols may inadvertently cause new, undesirable mutations in regenerated plants.

6.3 Methods

(For media recipes see Appendix D)

6.3.1 Plant material

Three-node cane cuttings (approximately 10mm in diameter) were collected from Pinot noir UCD5 vines (Lincoln University, Canterbury, New Zealand) during pruning. The cuttings were rooted in sand for one month and then potted in long-term potting mix (Appendix D).

Healthy apical growth tips (terminal 3-4cm of branches) were removed from Pinot noir UCD5 vines in the summer season (Feb 2011). Excised tips were washed for 30 seconds in 70% (v/v) ethanol, followed by 20 minutes of gentle agitation in 10% (v/v) bleach with 0.1% (v/v) Tween-20 and finally three rinses in sterile water. The terminal 2cm of each tip was removed with a scalpel and inserted into EG medium (Appendix D) in 10cm diameter tissue culture tubs for initiation of *in vitro* plantlets. Plantlets were incubated with an 8:16hr light:dark regime at 25°C with monthly subculture.

6.3.2 Preparation of embryogenic callus samples

Stage III inflorescences (DHEKNEY *et al.* 2009) were collected from Pinot noir clone UCD5 vines in an open vineyard at Lincoln University, Canterbury, New Zealand. The unopened flowers were sterilised by dipping entire inflorescences in a series of washes with gentle agitation: 70(v/v) ethanol for 30 seconds followed by 10% (v/v) bleach for 20 minutes and three washes with sterile water for 5 minutes each. Inflorescences were then stored at 4°C for two days.

Flower bud caps and anthers were removed from the pedicel by cutting through the base of the ovary. Anthers and stigma/style were placed on B medium (PERRIN *et al.* 2004) plates (approximately 100 anthers per plate), sealed with Parafilm and incubated at 25°C in the dark. After one month anthers showing callus formation were transferred to C_1^P medium plates (TORREGROSA 1998) for long-term maintenance with monthly sub culturing under the same growth conditions. Anthers that showed no callus formation but had not turned brown were moved to fresh B medium for another month. After the second month anthers producing callus were collected and the rest were discarded.

Liquid EC cultures were initiated according to Ben Amar and co-workers (2007). Briefly, 0.5g of EC from solid cultures was added to 50ml fresh medium (FM) with 2mg/ml larch wood arabinogalactan (Sigma-Aldrich) in a 500ml Erlenmeyer flask. Cultures were maintained on an orbital shaker at 150RPM, 25°C in the dark. Once per week half of the liquid medium was replaced with an equal volume of new FM. Over the first month, volume was gradually increased to 100ml by adding approximately 15ml extra medium at each transfer, after which it was maintained at 100ml.

At the time of stress exposure experiments, EC cultures had been maintained for approximately one year. Calli were still capable of regeneration into plantlets following transferral onto EG medium plates. Calli growing on solid medium were used two weeks after monthly subculturing. Liquid medium EC were used four days after weekly subculturing.

6.3.3 Stress treatments

Four tissue types were tested with a range of biotic and abiotic stresses. The tissue types used and the stress treatments tested on each are summarised in Table 6.1.

Table 6.1 Stress treatments tested on different tissue types. Asterisks (*) indicate a stress treatment applied to the specific tissue type.

Stress treatment	Solid-grown EC	Liquid-grown EC	*in vitro* plantlet leaves	Nursery vine leaves
Control	*	*	*	*
50mM NaCl	*	*		
100mM NaCl	*	*		
200mM NaCl		*		
4°C	*	*		
37°C	*	*		
UV	*	*		
Onozuka R-10	*	*		
acetosyringone	*	*		
live *Agrobacterium*	*	*	*	*
autoclaved *Agrobacterium*	*	*		
Botrytis extract	*	*	*	*
live *S. cerevisiae*	*	*	*	*
live *P. kluveri*	*	*		
live *A. pullulans*	*	*		
live *H. uvarum*	*	*	*	*
live *R. glutinis*	*	*		
live *C. magnus*	*	*		
autoclaved *S. cerevisiae*		*		
autoclaved *P. kluveri*		*		
autoclaved *A. pullulans*		*		
autoclaved *H. uvarum*		*		
autoclaved *R. glutinis*		*		
autoclaved *C. magnus*		*		

Biotic stressors

A crude *Botrytis cinerea* extract was prepared according to Holden (1950) from a wild isolate donated by Marlene Jaspers (Lincoln University, New Zealand). Briefly, damp bran was inoculated with the fungi and then incubated at 25°C for 6 days. The culture was then soaked in chloroform, dried at 37°C, and ground to a powder. Ten grams of powder was resuspended in 100ml water and left to stand for 1 hour. The suspension was passed through several layers of muslin cloth and

centrifuged to remove particulates. The final extract was dialysed with high-retention cellulose dialysis tubing (32mm flat width; Sigma-Aldrich) for 24 hours in 100mM phosphate buffer (pH7.0) at 4°C, during which time the buffer was changed thrice. Onozuka R-10 (Merck, Massachusetts, USA) was prepared by dissolving the powder form of the product in hormone-free liquid C_1^P medium to a concentration of 1mg/ml. Both the botrytis and Onozuka R-10 treatments were filter sterilised.

Fifty millilitre liquid cultures of *Agrobacterium tumefaciens* strain GV3101 were grown from single colony picks in LB broth (10g/l tryptone, 10g/l NaCl, 5g/l yeast extract, pH 7.5) for 48hrs at 28°C, 200RPM. The cultures were either used live, or autoclaved to kill the cells. Cultures were pelleted and resuspended to OD_{600} = 0.8 in hormone-free liquid C_1^P medium (HF-C_1^P) with 150µM acetosyringone. A control treatment of 150µM acetosyringone in HF-C_1^P was also prepared.

Six of the most common yeast species endemic to New Zealand vines (kindly supplied by Dr Matt Goddard, Auckland University) were also selected for treating somatic embryos. Live and autoclaved cultures of the yeasts were prepared in the same manner as the *A. tumefaciens* cultures, except that yeast were grown in YPD medium (10g/l yeast extract, 20g/l peptone, 20g/l glucose) for 48hrs at 22°C, 200RPM. Yeast culture treatments were also suspended to the same optical density in HF-C_1^P, but with no added acetosyringone.

Biotic stress exposure of EC involved dipping 0.5g of cells in 12ml HF-C_1^P medium containing the stressor in a 15ml falcon tube. Samples were shaken vigorously for 2 sec and incubated horizontally on a rotary shaker (100RPM) at room temperature for 8 minutes. Tubes were then placed vertically to allow cells to settle for a further 2 minutes. Liquid medium was removed with a pipette, and cells were spread evenly on HF-C_1^P-soaked filter papers on fresh C_1^P plates. Plates were incubated at 25°C in the dark for 48 hours. Before harvesting, EC were washed with HF-C_1^P on filter paper in a Buchner funnel until the wash medium ran clear to remove the majority of the microbial contamination.

Biotic stress exposure of grapevine leaf tissue involved adding 0.03% (v/v) Pulse surfactant (see Chapter 5) to the liquid treatments to improve tissue flooding. Young leaves (4-7cm diameter) of potted Pinot noir UCD5 vines in the nursery were treated *in situ* by dipping in liquid treatments for 10 minutes. Leaves were left on the vine to recover for 48 hours before collection. Young leaves (1.0-1.5cm diameter) of *in vitro* grown Sauvignon blanc UCD1 plantlets were excised from plantlets and submerged in liquid treatments. The leaves were then blotted dry on sterile filter paper and placed on MS plates (0.5X MS salts and vitamins, 30g/l sucrose, 7g/l phytoagar, pH 5.8) at 25°C to recover for 48hrs before collection.

Abiotic Stressors

Exposure of EC to abiotic stresses involved first transferring 0.5g of cells onto HF-C_1^P-soaked filter papers on fresh C_1^P plates. For temperature shock treatments, plates were placed at either 4°C or 37°C for 48hrs, after time the cells were harvested, as for the biotic stress exposure treatments. For UV shock treatment, cells were exposed to short-wave UV-C radiation (11.2 UV Watts at 254nm for 10 minutes) and then allowed to recover for 48hrs under normal growth conditions. Salt stress was applied by transferring 0.5g of EC onto fresh C_1^P plates supplemented with NaCl at a final concentration of 50mM, 100mM or 200mM.

6.3.4 Sample preparation

Tissue collection involved snap freezing the samples in liquid N_2 and then grinding them with a mortar and pestle in the presence of liquid N_2. RNA extractions were performed on 100mg of ground tissue using the Spectrum™ Plant Total RNA Kit (Sigma-Aldrich) according to the manufacturer's instructions. On column DNAse treatment was performed during the extraction using the On-Column DNase I Digestion Set (Sigma-Aldrich) according to the manufacturer's instructions, except that incubation times were doubled. To control for qPCR amplification of RNA contamination from yeast cells, RNA was extracted from pure cultures of each of the yeast strains using the same protocol as for the plant cells. These were processed in parallel with the plant RNA.

RNA yields were quantified using a Qubit® fluorometer (Life Technologies) and assayed for purity with a NanoDrop® 1000 UV-vis spectrophotometer (Thermo Fischer Scientific). Each sample (300ng) was run on a denaturing formaldehyde gel to check the sample integrity. To check for DNA contamination, one microliter of each extraction was used as template for a PCR reaction with actin qPCR primers (Appendix C). In addition to the template RNA, the reaction mix contained 0.5U Taq polymerase (Qiagen), 1X PCR buffer, 200µM each of dNTPs and 0.2µM of each primer. Reaction conditions were: one cycle at 95°C for 1 minute; 35 cycles of 95°C for 30s, 55°C for 30s, 72°C for 1 minute; one cycle at 72°C for 5 minutes. Samples that produced amplification products were treated with Turbo DNA-free (Ambion) according to the manufacturer's protocol and re-tested.

For first strand cDNA synthesis, 300ng of RNA were used as template in a 10µl reaction containing 0.5µl BluePrint Enzyme Mix (Takara Bio Inc.), 5µM of random hexamers, 2.5µM of oligo(dT) primer and 1X BluePrint Buffer. The reaction was incubated at 37°C for 15 minutes, followed by reverse transcriptase inactivation at 85°C for 5 sec. Reactions were diluted 30-fold by the addition of 290µl RNase-free H_2O.

6.3.5 Retrotransposon transcript quantification

Primer design

Primers sequences for the grapevine reference genes ACTIN, EF1α, GAPDH and SAND recommended by Reid and colleagues (2006) were used. Retrotransposon-specific qPCR primers were designed to flank the 5' boundary of the 3'LTR so that only transcripts from complete retrotransposon insertions are amplified (see Figure 6.1). Primer sequences are given in Appendix C.

Figure 6.1 Location of qPCR primer binding sites (red arrows) on a complete retrotransposon transcript. Figure not drawn to scale.

Preparation of standards

Standards were produced by cloning fragments amplified by PCR using each qPCR primer pair. For retrotransposon standards Pinot noir UCD5 genomic DNA was used as a template and for reference gene standards Pinot noir UCD5 young leaf cDNA was used as a template. The amplification mix contained 2.5 units of PrimeSTAR HS DNA polymerase, 200µM each of dNTPs, 0.2µM of each primer, 20ng template DNA and 1X PrimeSTAR buffer in a 50µl volume. Reaction conditions were: 95°C for 1 min, 35 cycles of 98°C for 10 sec, 60°C for 5 sec, 72°C for 1 min. The amplification products were visualised on a 3.5% (w/v) agarose TBE gel (Figure 6.4), purified using the AxyPrep Gel purification kit (Axygen) and cloned into the pGEM-T Easy vector system (Promega) using the manufacturer's Rapid Ligation protocol. Briefly, 50ng vector and 10ng PCR product (1:3 molar ratio) were combined with 3 Weiss units of T4 DNA ligase (Promega) and 5µl 2X Rapid Ligation Buffer (Promega) in a 10µl reaction. The reaction was incubated at room temperature for 1 hour and then 2µl of ligation mix was used to transform chemically competent DH5α *E. coli* cells according to Sambrook and Russell (2001) Appendix E.1).

Since the multiple cloning site (MCS) of the pGEM-T Easy vector system lies within the *lacZα* gene, blue/white screening was used to select *E. coli* colonies with insert-containing plasmids. Inserts were checked by colony PCR (Appendix E.4) using the relevant qPCR forward and reverse primers (Appendix C). Colonies which produced amplification products of expected size were transferred to 3ml LB broth with Ampicillin (50mg/l) and grown overnight at 37°C, 250RPM. Plasmid DNA was purified from each of these cultures using the miniprep plasmid extraction technique as described by Sambrook and Russell (2001), checked by Sanger sequencing (Appendix E.5) using the vector-specific

T7 promoter primer and then linearised enzymatically with *Nco*I (New England Biolabs). The AxyPrep PCR Clean-up Kit (Axygen) was used to clean linearised plasmid DNA, which was then quantified with a Qubit fluorometer (Life Technologies).

qPCR reactions

Real-time quantitative PCR reactions were performed using the Eco qPCR system (Illumina). For each gene analysed, a 10-fold dilution series of qPCR standards ranging from 10^{-1} to 10^{-8}ng/µl was prepared. Each experiment involved triplicate reactions for each sample, duplicate reactions for each of five standard dilutions (10^{-2}, 10^{-4}, 10^{-5}, 10^{-6} and 10^{-8}ng/µl), and duplicate non-template control reactions per gene tested.

Preparation of standards and loading of samples and master mix into the run plate were performed using an epMotion 5070 liquid handling robot (Eppendorf) to minimise pipetting variation. Reactions were performed in 10µl volumes containing 0.2mM of each primer, 3µl of diluted cDNA and 5µl SYBR Premix (Takara). Two-step thermocycling conditions included and initial denaturation at 95 °C for 5 min, 40 cycles of 95°C for 30 s and 60°C for 30 s, followed by 95°C for 15 s and real-time melt curve analysis from 55°C to 95°C with fluorescence recorded every 0.3°C. A variation of 0.25 C_T was used as the threshold of acceptable standard deviation among technical replicates.

Reference gene comparisons

The suitability of four reference genes selected for normalisation of transcript data was tested by running qPCR reactions with each of the genes on nine samples representing three treatment groups (control, live *A. tumefaciens* and 37°C heat stress). The geNorm algorithm (VANDESOMPELE *et al.* 2002) was used to rank reference gene expression stability based on pairwise C_T comparisons and to determine the optimal number of reference genes for use in normalisation of transcript data.

Transcript quantifications

Specific amplification efficiencies were determined for each primer set from the efficiency curves plotted using the standard reactions. Quantification was calculated according to the Pfaffl method from the mean of the technical replicates (PFAFFL 2006). In each experiment, transposon expression was normalised to the geometric mean of the expression of two reference genes (Actin and EF1α). Expression changes were then calculated relative to the average of the three control samples.

6.3.6 *In vivo* LTR-reporter assays

Vector construction

For each of four retrotransposon families (*Gret1*, *Edel*, *Cremant* and *Noble*) a canonical element that showed high integrity, high LTR identity and a complete open reading frame was selected from the

PN40024 genome sequence data. Primers were designed to amplify these sequences from within 300bp upstream of the retrotransposon insertion to directly before the first internal ORF, a region that spans the 5' LTR, PBS and 5' untranslated region (Figure 6.2). Primers were also designed to amplify the *cauliflower mosaic virus* 35S promoter (CaMV 35S) from the pART7 plasmid (GLEAVE 1992) for use as a constitutive control for promoter analysis (Appendix C).

Figure 6.2 Primer binding sites for amplification of retrotransposon 5' LTR + 5' UTR fragments.

Amplification products were gel purified and cloned into the pENTR/D-TOPO vector system (Invitrogen) according to the manufacturer's protocol. Plasmids were transformed into chemically competent *E. coli* DH5α cells (Appendix E.2) and grown with Kanamycin (50mg/l) selection. Plasmid extraction was performed as described for qPCR plasmids above and inserts were checked by Sanger sequencing using vector-specific M13 primers (Appendix E.5). Inserts were then transferred into the Gateway insertion site directly upstream of GFP-GUS fusion gene in the expression vector pBGWFS7 (KARIMI *et al.* 2002) by LR Clonase recombination reaction (Invitrogen) according to the manufacturer's protocol. In addition to the entry vectors described above, a recombination reaction was performed between pBGWFS7 and a pENTR plasmid containing no insert to produce a promoter-free control vector.

Recombination products were once again transformed into chemically competent *E. coli* DH5α cells and grown with Spectinomycin (50mg/l) selection. Plasmids extracted from single colony cultures (described above) were checked for correct insert by restriction digestion with 5U of *Eco*RV (New England Biolabs). All plasmids gave the predicted digestion pattern. The expression vectors were named pB:Gret1-GG, pB:EdelLU-GG, pB:CremantLU-GG, pB:NobleLU-GG, pB:Empty-GG and pB:35S-GG. The vectors were transformed into *A. tumefaciens* by electroporation (Appendix E.3) and grown on LB plates with spectinomycin (50mg/l) selection. Diagrams for each of the vectors are given in Appendix A.

An *A. tumefaciens* strain containing the vector pB:35S-3GFP was used as a transformation control. This vector was produced by inserting the 35S-MCS *Not*I fragment from pART7 into a the *Not*I site of pMLBart, a pART27-derived plasmid containing the Bar gene in place of NptII. A Gateway cassette was then inserted into the *Hind*III site of the MCS to form pMLBart:35S-GW. An LR Clonase reaction performed (Invitrogen) between the this plasmid and a pENT1A plasmid containing a triple GFP

coding sequence was used to introduce the triple GFP ORF behind the 35S promoter in the new plasmid pB:35S-3GFP.

Tobacco transient infiltrations

Agrobacterium tumefaciens cultures containing each of the above vectors were grown for 2 days in 10ml LB broth with spectinomycin (50mg/l) at 28°C, 200RPM. The bacteria were pelleted by centrifugation and resuspend in 40ml MS liquid medium with 150µM acetosyringone (Appendix D). After 2hrs of incubation at 28°C, 200RPM, the cell density was adjusted to OD_{600nm} = 0.2 with additional liquid medium. One millilitre syringes were used to infiltrate *N. tabaccum* leaf tissue by gently pressing suspension cultures into the underside of young leaves (2-3cm across) on potted 3-month old plants. After 48hrs at normal growth conditions, infiltrated leaves were harvested and GUS staining assays were performed as described in Appendix E.6.

Stable LTR-reporter tobacco plants

Young leaves of sterile *N. tabaccum* plantlets in tissue culture (approximately 2cm across) were infiltrated *A. tumefaciens* strains carrying each of the vectors described above. After 48hrs under normal growth conditions (25 °C, 16hr:8hr light:dark cycle) leaves were removed and placed in tissue culture tubs containing 50ml MS regeneration medium (Appendix D) with 20mg/l DL-phosphinothricin (glufosinate ammonium salt) and 200mg/l cefotaxime. When shoots began to appear (after 4-5 weeks), these were removed and transferred to MS rooting medium with the same phosphinothricin and cefotaxime concentrations.

After rooting, T_0 plantlets were transferred to Black Magic potting soil (The Scotts Company LLC) in 10cm x 10cm x 15cm pots and maintained until flowering. Unopened flowers were bagged to force self-fertilisation. The T_1 seeds collected from the pods of these flowers were kept at 4°C for two days and then sown in the same small pots containing potting mix covered with a thin layer of vermiculite. Seeds were watered with 20mg/l DL-phosphinothricin to select for transgenic constructs. Self-crossing, collection of T_2 seed and sowing of new seed with phosphinothricin selection were the same as for the T_1 seed.

At each generation, DNA was collected by crushing the leaves of each plant against FTA cards (Whatman). A single 1mm punch was taken from each card and washed with FTA reagent (Whatman) according to the manufacturer's recommendations. Transgene presence was confirmed by direct PCR on the cleaned FTA punch with the forward primer used to amplify the LTR from genomic DNA and the GFP border check reverse primer (Appendix C). Products were checked on a 1% (w/v) agarose TBE gel by electrophoresis at 100V for 40 minutes.

Once T_2 plants had grown to approximately 10cm in height, suspensions of *S. cerevisiae* and *H. uvarum* (cells resuspended to OD600nm = 1.0 in liquid MS medium) were infiltrated into the underside of leaves of three plants per transgenic line using a 1ml syringe. The leaves of three plants of each line were also pressed with an empty syringe to reproduce the same wound sites as those arising from infiltration. After two days these leaves were removed and stained for GUS expression assays.

Transient transformation on grapevine leaves

Transient *Agrobacterium*-mediated transformation of young leaves of *in vitro* grown S. blanc plantlets with LTR-reporter vectors was performed according to the protocol developed (see Chapter 5). Five leaf disks were used for transformation with each expression vector. After dipping, leaf discs were co-cultivated with *A.* tumefaciens on MS plates (Appendix D) for two days before analysis GUS expression assay (Appendix E.6).

6.4 Results

6.4.1 Initiation and maintenance of embryogenic callus cultures

Although efficiencies were low, anthers proved to be effective explant material for the initiation of EC cultures. In total, 2,477 Pinot noir UCD5 anthers were plated on B medium. From these, 72 EC masses were collected (2.90% initiation efficiency). All EC grew and proliferated well after transfer to C_1^P medium. Transferring EC onto MPM01 medium for one month prior to long-term maintenance on C_1^P medium made no difference to viability and proliferation of the cells. Calli did not proliferate well on MPM1 medium recommended by Perrin and colleagues (2001) and were therefore maintained on C_1^P medium plates.

6.4.2 Sample processing

To test the TE transcriptional changes in these EC cells, the cultures were exposed to a range of stress treatments. Embryogenic callus samples treated with UV-C light showed browning 48 hours after treatment. Cells treated with botrytis extract, live *A. tumefaciens* and Onozuka R-10 also showed slight browning. No colour change was see in other treated cultures. Yeast growth was visible on EC co-cultivated with the live yeast cultures, although this was easily removed after several rounds of rinsing in a Buchner funnel. After yeast were washed off, these EC showed a cream colour comparable to the control samples.

Before processing for qPCR, the integrity of RNA extracted from the stressed EC cultures for transcript analysis was checked by gel electrophoresis. Denaturing agarose gels demonstrated typical patterns of high integrity RNA for all extractions (two distinct ribosomal RNA bands) except for the EC samples co-cultivated with live yeast strains for 48hrs (Figure 6.3). In these samples the major ribosomal RNA bands were visible although less pronounced and other distinct bands (possibly yeast rRNA) were sometimes present. Washing somatic embryos with an RNase inhibitor (RNAlater, Invitrogen) before performing extractions had no effect on the RNA integrity.

To test for DNA contamination, all RNA samples were used as templates for PCR amplification. Contaminating DNA was identified in a low number of samples (<10%). This contamination was removed by treating samples with Turbo DNA-free (Ambion) and the samples were re-tested to confirm this (see 6.3.4).

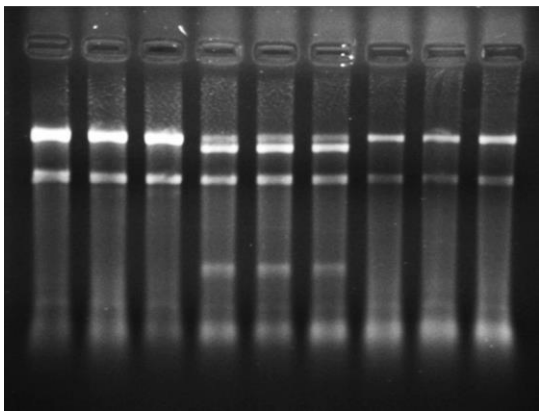


Figure 6.3 RNA integrity following stress treatments. Moderate RNA degradation was seen in samples treated with live yeast cultures. All other stress treatments had not distinguishable effect on RNA integrity. Lanes 1-3: Control EC, lanes 4-6: EC stressed with *S. cerevisiae*, lanes 7-9: EC stressed with *H. uvarum*. Image is of a 3.5% (w/v) agarose formaldehyde gel loaded with 500ng of each RNA sample and run at 80V for 1 hour.

6.4.3 Retrotransposon transcript quantification

Experimental design for quantitative PCR

Each of the primer pairs designed to amplify TE transcript targets produced distinct amplification products (Figure 6.4) when amplifications were performed on genomic DNA. These amplification products were effectively cloned for use as standards in qPCR assays.

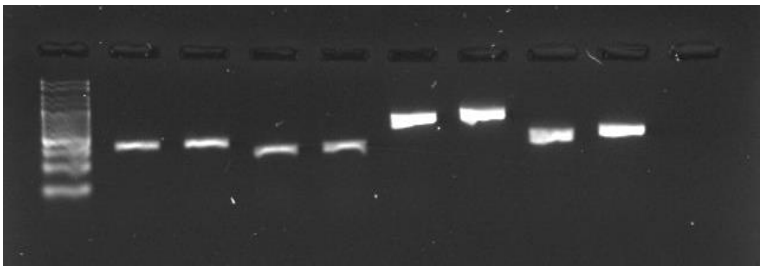


Figure 6.4 Amplification products of the qPCR primer sets. Lane 1: 25bp ladder, lane 2: Actin, lane 3: EF1α, lane 4: SAND, lane 5: GAPDH, lane 6: *Gret1*, lane 7: *Edel*, lane 8: *Cremant*, lane 9: *Noble*, lane 10: negative control. Primer pairs for qPCR amplification are given in Appendix C.

To check the suitability of the reference genes, qPCR amplifications were performed on a subset of samples using all four of the potential reference genes as targets (see 6.3.5). Each of the potential reference genes quantified showed low variation across the samples analysed (geNorm M value < 0.5). Analysis the data using geNorm showed that the two most stable reference genes, Actin and EF1α, were satisfactory for data normalisation (Figure 6.5). When run alongside test samples, the standard curves plotted for each gene produced R^2 values of greater than 0.99 and show amplification efficiencies between 95% and 110% (Figure 6.6).

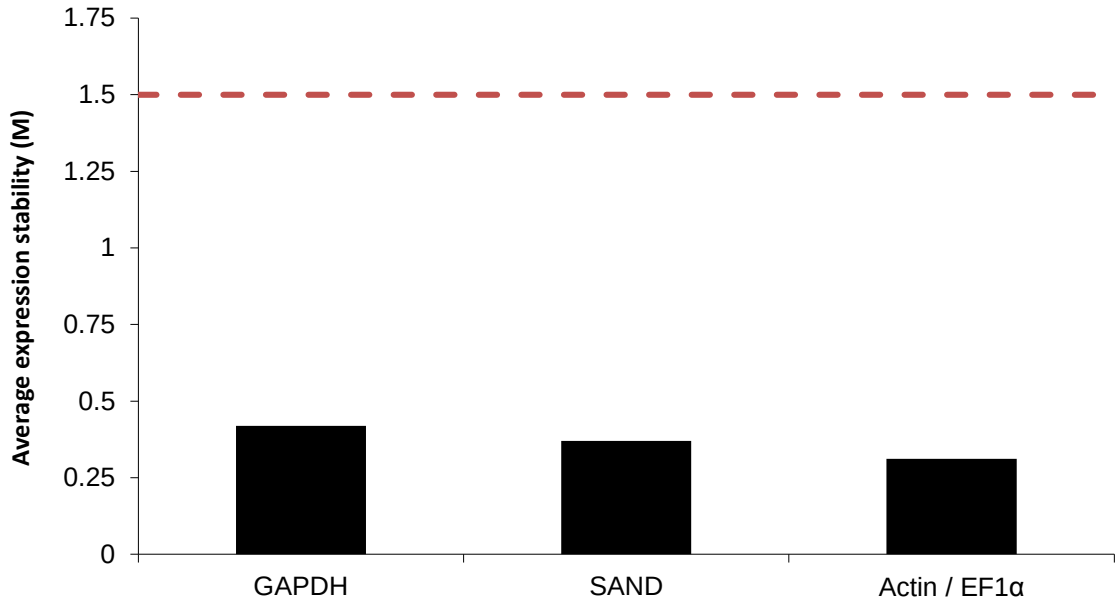


Figure 6.5 Relative stability of potential reference gene transcripts. Higher M-values (y-axis) represent increased variation among samples. The x-axis lists tested genes in order of their stability (decreasing M-value). Since the geNorm algorithm computes the relative stability of genes, the two most stable genes (Actin and EF1α in this case) cannot be discriminated. The dashed red line represents the threshold of maximum variation recommended for reference gene selection.

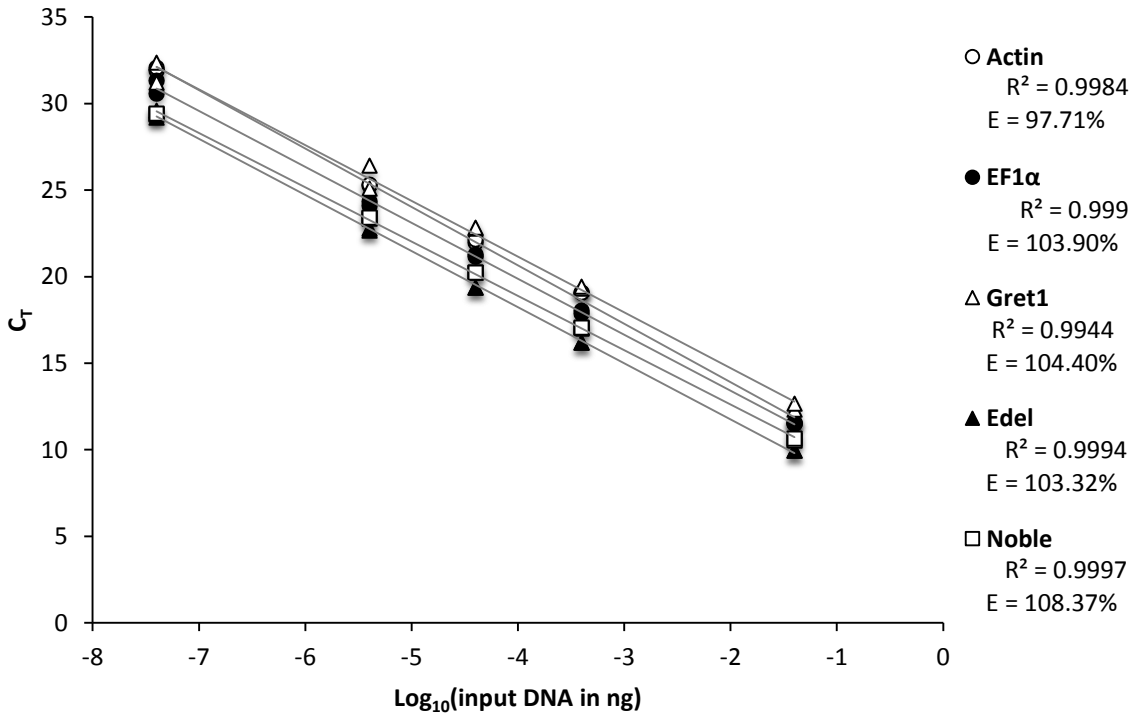


Figure 6.6 Reference gene standard curves. Produced from the real-time amplification of a linearised plasmid dilution series (R^2: coefficient of determination for the standard curve, E: efficiency).

Retrotransposon expression changes in solid-medium EC culture

Retrotransposon transcription was first tested in EC cultures grown and stress-tested on solid medium. For those samples in which retrotransposon transcription was increased, the magnitude of this effect varied greatly between callus samples (Figure 6.7). In order to display both the variation between samples and the fold change over time per sample, replicates are shown separately in the graphs.

Embryogenic calli that had been growing on solid medium showed little variation in retrotransposon expression after 48 hours of exposure to abiotic stressors, Onozuka R-10, botrytis extract or *Agrobacterium*. In contrast, those samples that had been incubated with live yeast cultures showed large increases in retrotransposon transcript levels relative to the reference genes (Figure 6.7). *Gret1* transcripts displayed the greatest increase, with *Edel* and *Noble* showing lower levels of increase.

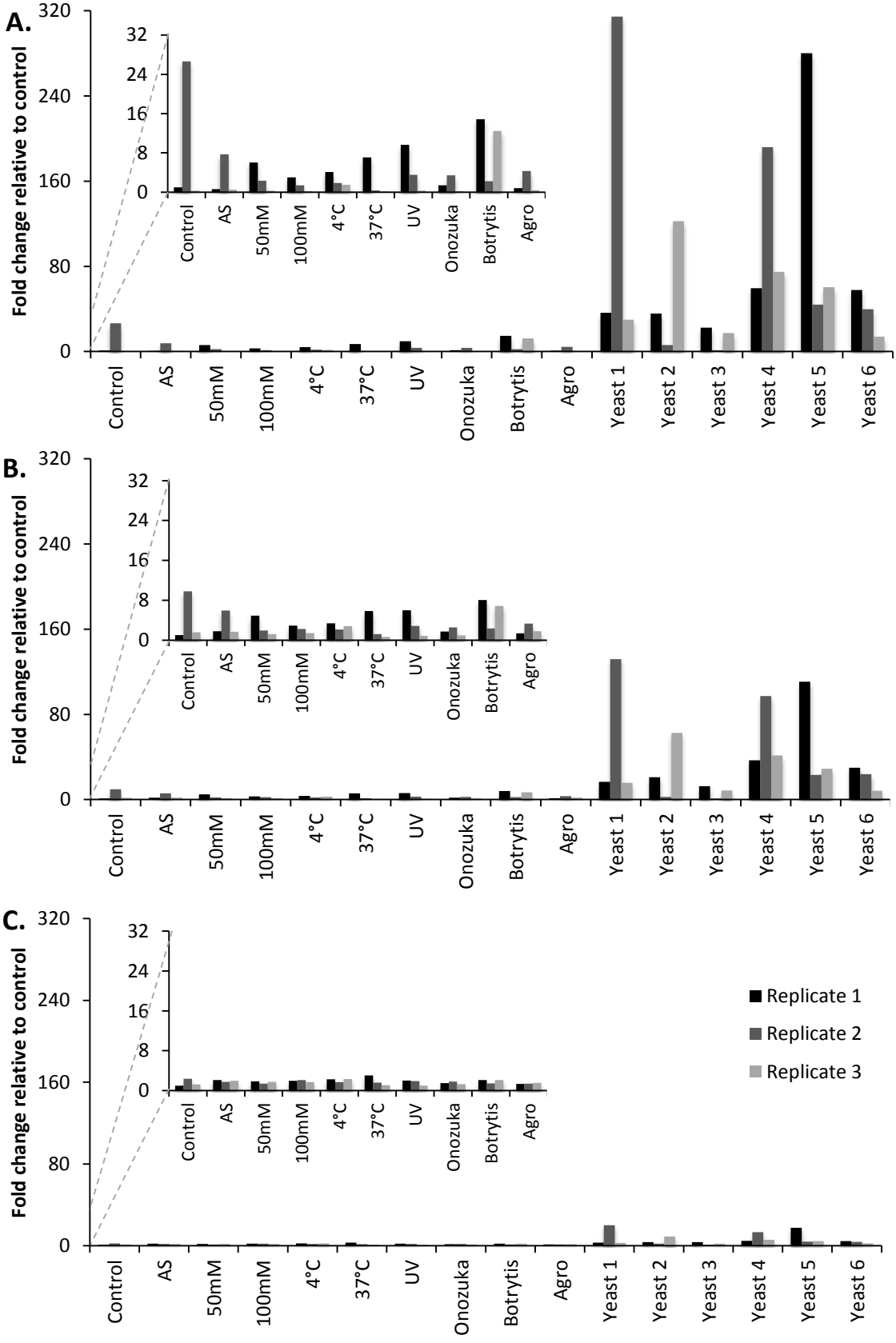


Figure 6.7 Retrotransposon expression changes in EC from solid-medium cultures exposed to stress. A: *Gret1*, **B**: *Edel*, **C**: *Noble*. Transcript abundance was measured 48 hours after stress exposure was initiated. Normalisation was performed using the geometric mean of Actin and EF1α expression and expression is shown relative to control replicate 1. Biological replicates are shown separately to illustrate the variation between callus samples as well as the relative expression change. Inset graphs show less responsive samples on a 10-fold smaller scale (Yeast 1: *S. cerevisiae*,

Yeast 2: *P. kluveri*, Yeast 3: *A. pullulans*, Yeast 4: *H. uvarum*, Yeast 5: *R. glutinis*, Yeast 6: *C. magnus,* AS: Acetosyringone, Agro: *A. tumefaciens, 50mM: 50mM NaCl stress, 100mM: 100mM NaCl stress*).

Comparison of the relative expression change between the two reference genes, Actin and EF1α showed that the high variation between biological replicates seen for the retrotransposon transcripts was not present in the reference genes (Figure 6.8).

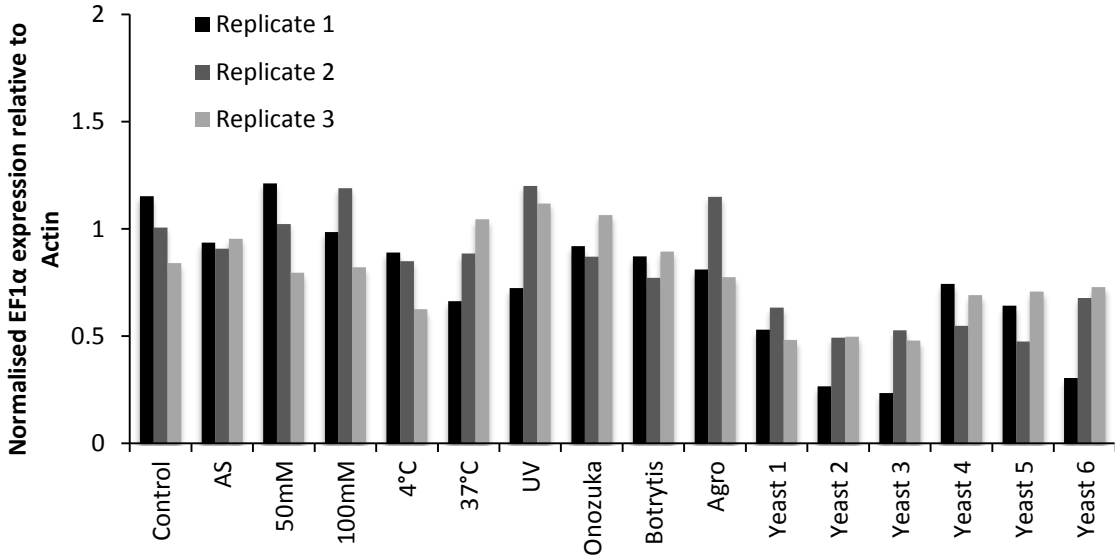


Figure 6.8 Expression levels of EF1α relative to Actin across stressed EC samples. Biological replicates are shown separately to illustrate the variation between callus samples as well as the relative expression change (Yeast 1: *S. cerevisiae*, Yeast 2: *P. kluveri*, Yeast 3: *A. pullulans*, Yeast 4: *H. uvarum*, Yeast 5: *R. glutinis*, Yeast 6: *C. magnus, 50mM: 50mM NaCl stress, 100mM: 100mM NaCl stress*).

Visual inspection of EC proliferating on solid medium showed dissimilarities between colour, friability and morphology among calli originally propagated from the same mass (Figure 6.9). This morphological plasticity becomes evident after about 2 weeks of cultivation on C1P medium. It appears that although reference gene transcript levels are unaffected, stimulation of retrotransposon expression is highly variable between callus samples.

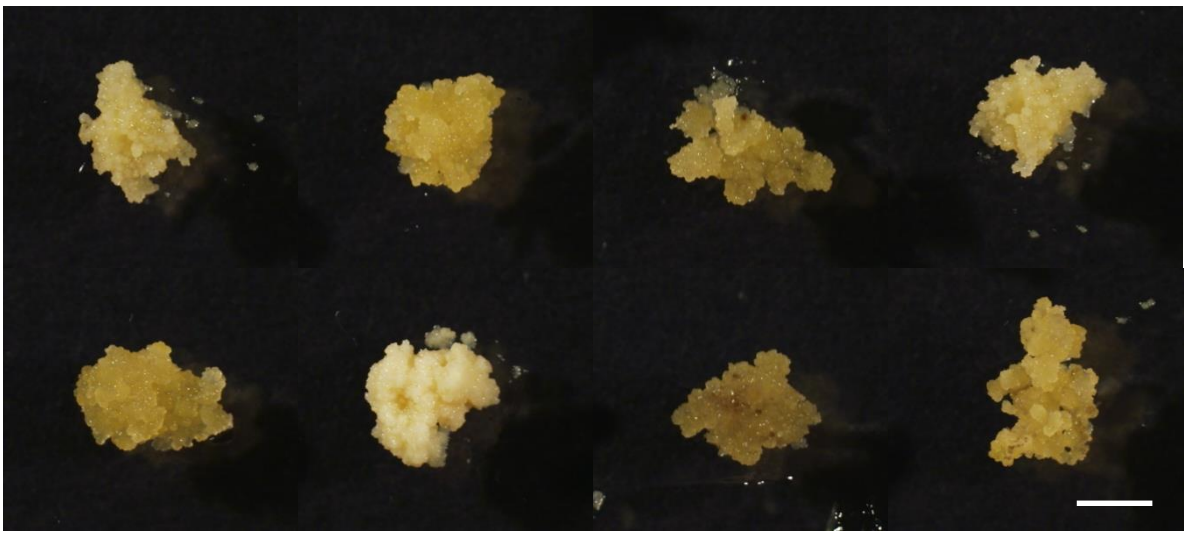


Figure 6.9 Morphological variation among embryogenic calli. All calli were propagated from a single Pinot noir UCD5 embryogenic callus mass. Scale bar = 3mm.

Callus samples treated with yeast cultures were rinsed well to remove the yeast cells before RNA extraction. In order to test whether the high retrotransposon expression in these samples could be attributed to contamination of yeast RNA, cDNA synthesis and real-time qPCR reactions were performed on purified RNA extracted from each of the yeast cultures according to the same protocol. If yeast RNA were contaminating EC qPCR data, the single observed melt peak would be the same in yeast and EC-stressed samples, and amplification would be much sooner when purified yeast RNA was used as a template.

In contrast, these samples showed amplification much later than embryogenic callus samples (4-9 C_T delay) and melt curve analysis revealed different amplification curves. Such results are typical of non-specific primer binding. Yeast-stressed EC showed a melt peak identical to all other EC samples, but different to yeast RNA alone. This indicates that RNA from any remaining yeast not removed during the wash step did not affect to the expression data.

Retrotransposon expression changes in liquid EC cultures

In an attempt to improve uniformity of biological replicates, EC cells grown in flasks of liquid medium were also subjected to stress treatments and analysed by qPCR. The mixture of calli in these cultures was expected to result in higher biological similarity between replicate samples collected. Autoclaved samples of each of the yeast strains were also prepared to test whether the retrotransposon expression increase seen could only be induced by live yeast. Due to the increased number of samples, the relative expression of two only retrotransposon families (*Edel* and *Noble*) were initially compared to two reference genes (Actin and EF1α), to test whether these cultures were more appropriate than EC grown on solid medium for testing retrotransposon expression.

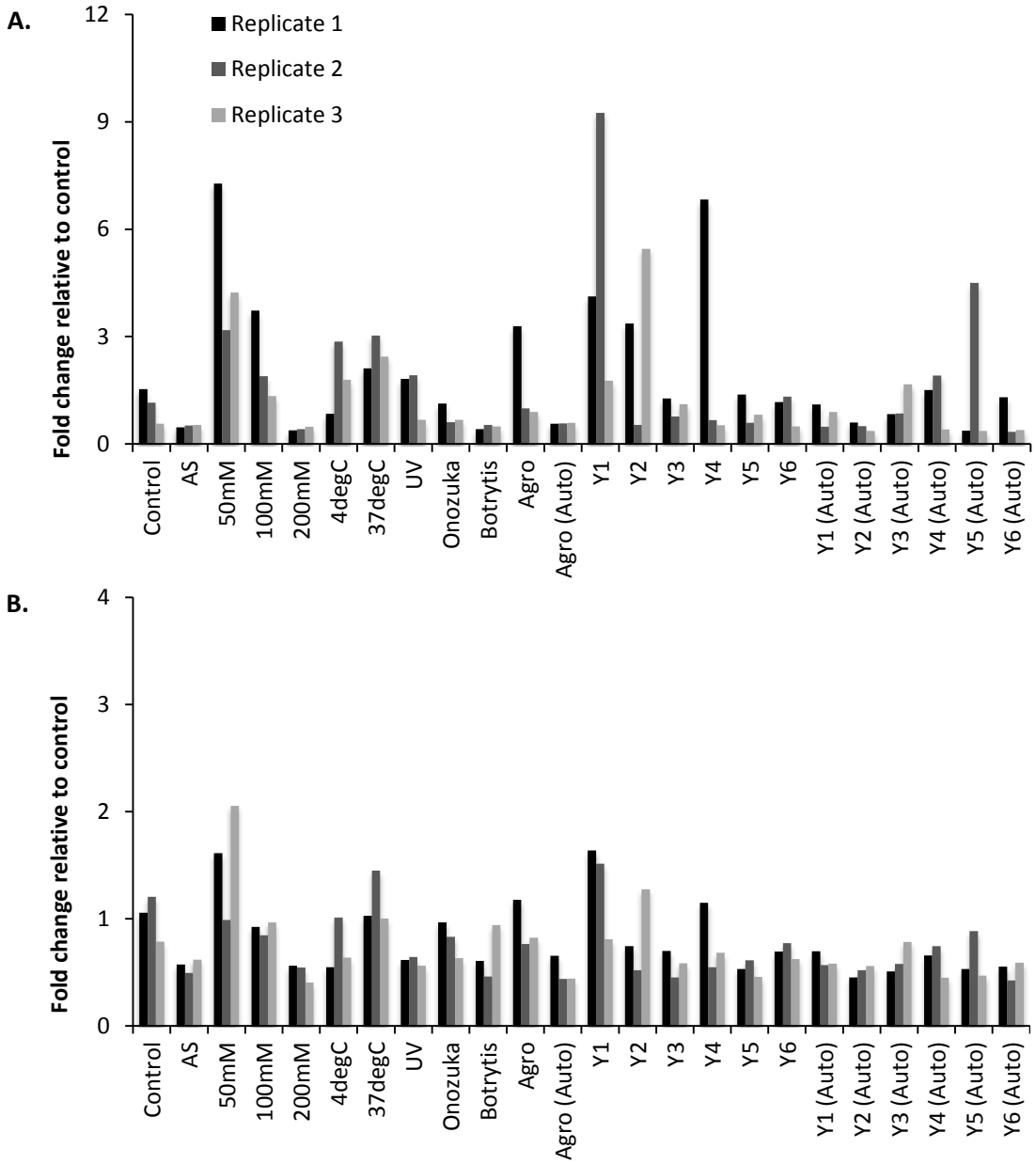


Figure 6.10 Retrotransposon expression changes in liquid-culture EC exposed to stress.
A: *Edel*, **B:** *Noble*. Transcript abundance was normalised to the geometric mean of Actin and EF1α expression, and is shown relative to the average of 3 replicate control (untreated) samples. Biological replicates are shown separately to illustrate the variation between callus samples as well as the relative expression change (Y1: *S. cerevisiae*, Y2: *P. kluveri*, Y3: *A. pullulans*, Y4: *H. uvarum*, Y5: *R. glutinis*, Y6: *C. magnus*, Auto: autoclaved, AS: Acetosyringone, Agro: *A. tumefaciens, 50mM: 50mM NaCl stress, 100mM: 100mM NaCl stress*).

Retrotransposon transcript levels of biological replicates show significantly less variation than was the case with calli grown on solid medium (Figure 6.10). The low-level (3 to 6-fold) increases in *Edel* expression seen in some of the abiotically stressed samples are similar to the changes seen in solid-medium grown EC following the same stresses. However, the very high degree of induction seen following co-cultivation with live yeast is absent in these samples.

Retrotransposon expression changes in leaf tissue

To test whether the TE response was specific to EC tissue, the change in TE transcription was compared between leaf tissue and EC cultures exposed to biotic stressors. No significant change in retrotransposon expression was seen in the case of any of the leaf samples. In contrast, the EC cultures once again showed the same uneven but high retrotransposon transcript numbers following cultivation with yeast strains.

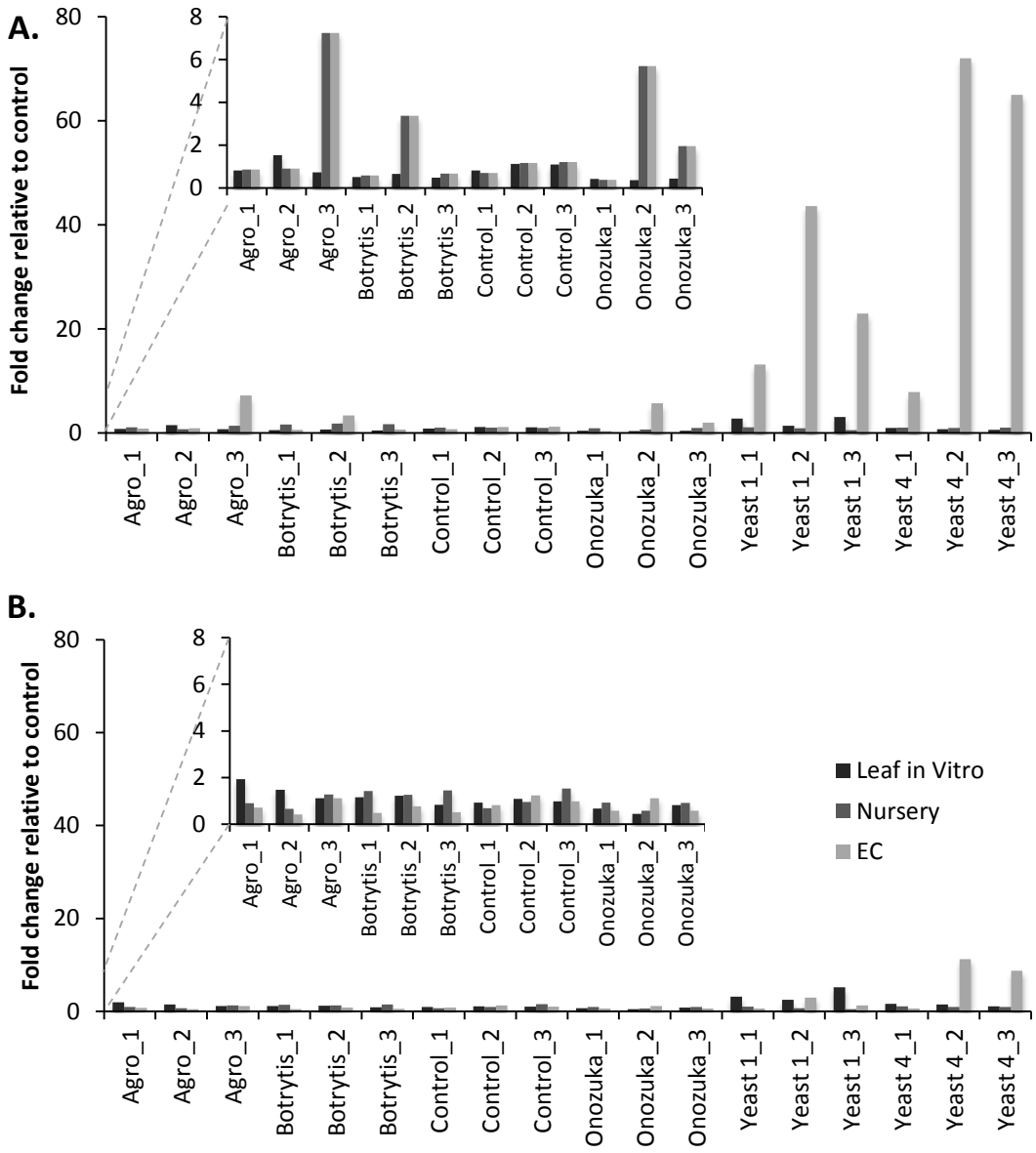


Figure 6.11 Expression of TEs after shock treatments in three different explant types. A: *Gret1*, **B:** *Edel*. (Leaf *in vitro*: leaf of S. blanc tissue culture plantlet, Nursery: leaf of potted 2-year old P. noir vine, EC: embryogenic callus). Transcript abundance was normalised to the geometric mean of Actin and EF1α expression, and is shown relative to the average of 3 replicate control (untreated) samples. Inset graphs show less responsive samples on a 10-fold smaller scale (Yeast 1: *S. cerevisiae*, Yeast 4: *H. uvarum*).

To determine the abundance of transcripts from each retrotransposon family, absolute quantifications were analysed by comparing samples against standard curves of known template amount. These were normalised against both reference genes in order to account for RNA quantification error and variations in reverse transcription reaction efficiency. Results from this analysis (Figure 6.12) showed that *Cremant* transcript is close to the limits of detection ($<0.3x10^{-5}$ pg/reaction) in control samples, but increases to 1-3% of the geometric mean of the reference genes after EC is exposed to *S. cerevisiae*. The Noble TE family, in comparison, has a high background transcript level (around 25% that of the reference gene mean) and in response to the yeast stress the transcription levels increase greatly, to a level above that of the reference genes in some cases.

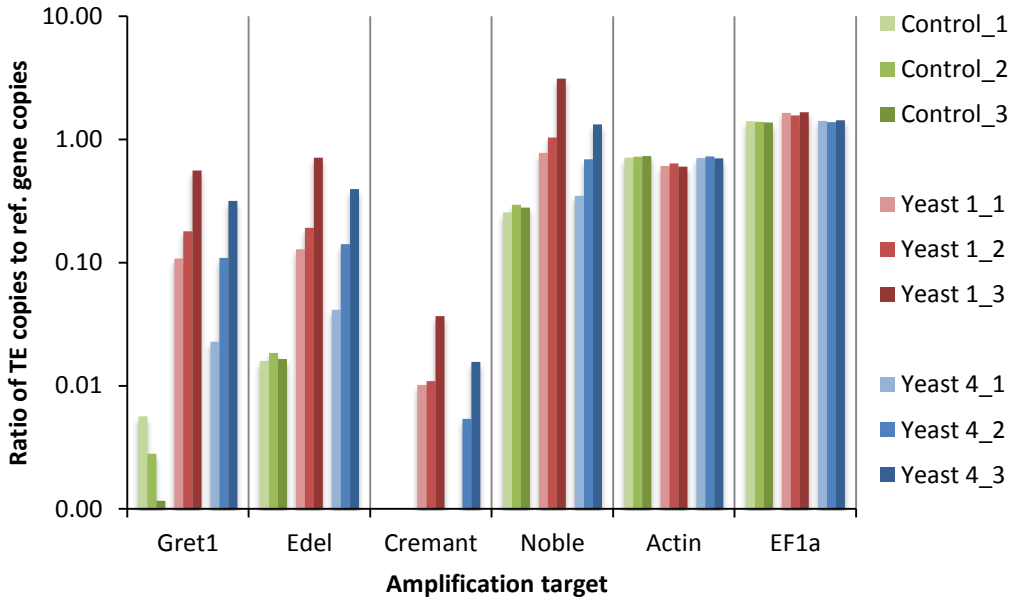


Figure 6.12 Absolute transcript abundance normalised to reference genes. Absolute transcript abundance was calculated using the standard curve method and then normalised to the geometric mean of Actin and EF1α. Triplicate biological samples are plotted separately to show variability among biological replicates. Transcript abundance for each reference gene is included for comparison (Yeast 1: *S. cerevisiae*, Yeast 4: *H. uvarum*).

When normalised transcript abundance is viewed relative to controls samples (Figure 6.13), the fold change of Cremant is far higher than that of Noble. This is clearly because of the background transcription rate of Cremant in untreated EC cultures is so much lower than that of Noble.

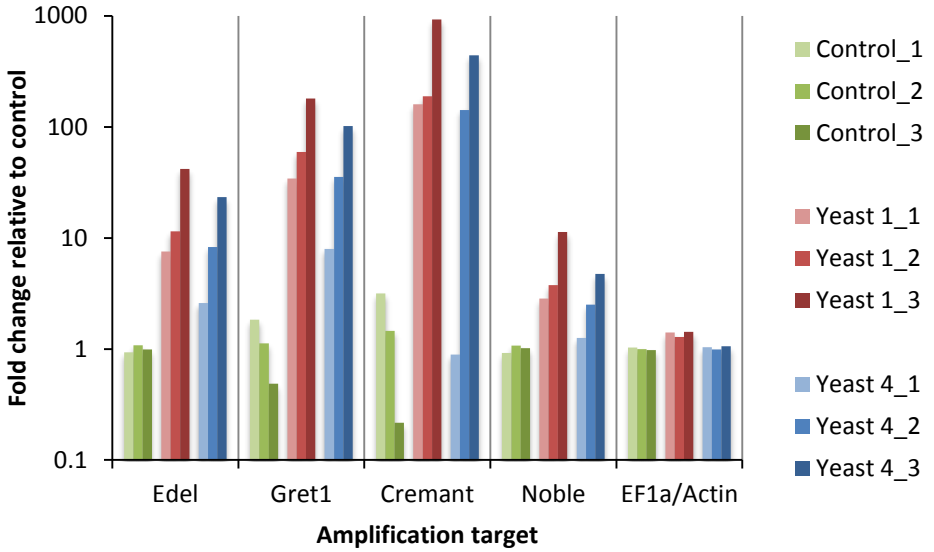


Figure 6.13 Relative expression of EC from solid medium following yeast exposure. TE transcripts have been normalised to both reference genes and are plotted on a logarithmic scale. Triplicate biological samples are plotted separately to show variation. The EF1α transcript levels are plotted relative to actin to show the stability of these genes for comparison (Yeast 1: *S. cerevisiae*, Yeast 4: *H. uvarum*).

6.4.4 The promoter activity of grapevine TE sequences

Tobacco transformation assays

The transcriptional activity of grapevine TE sequences was tested in tobacco using promoter-reporter constructs. Transient transformation assays were used to test the ability of the TE LTR sequences to promote expression in a model plant species, tobacco. Infiltration of tobacco leaves with *A. tumefaciens* carrying the pB:Empty-GG negative control vector resulted in neither fluorescence nor blue colour following GUS staining. Transformations with the pB:35S-3GFP positive control vector produced strong fluorescence at the infiltration sites under UV light, indicating high transformation efficiency. Transformation with pB:35S-GG positive control vector resulted in a strong blue signal following GUS staining of the tissue.

A fairly low level fluorescence was visible at the infiltration sites of each of the LTR-reporter-carrying *A. tumefaciens* cultures. The variation between constructs was more obvious after GUS staining, with *Noble* and *Cremant* LTR sequences driving strong expression of the GUS gene, the Edel LTR sequence inducing moderate expression and no expression observed from the pB:*Gret1*LU-GG vector (Figure 6.14).

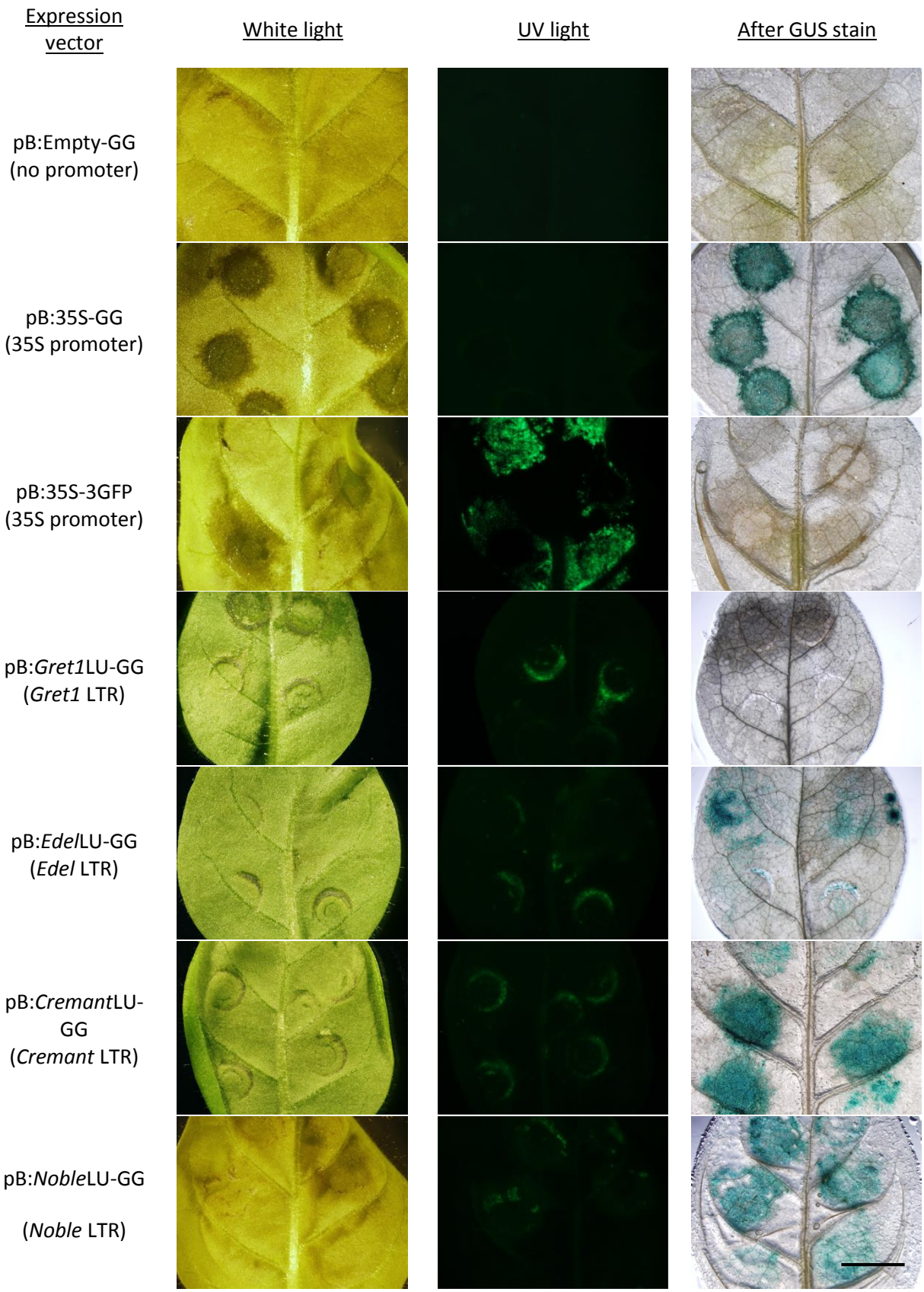


Figure 6.14 Transient *Agrobacterium* infiltration of tobacco leaves with LTR-reporter constructs.

Expression vectors either had no promoter . Each image is representative of three biological replicates. Scale bar = 5mm.

Stable transgenic tobacco plants were used to test the background expression levels from grapevine TE LTRs, and to see whether the expression is altered by infiltration with yeast species, a stress shown to increase the transcription of these TEs in grapevine (see 6.4.3). Infiltration of the leaves of transgenic T_2 *N. tabaccum* plants with yeast cultures resulted in no detectable difference to wounding alone. Plants carrying the promoter-free control construct and those with *Gret1* or *Edel* LTR sequences preceding the GUS reporter gene showed no expression under any conditions. The *Cremant* LTR sequences produced high levels of reporter gene expression, particularly in the leaf veins. Expression appeared to be slightly increased in living tissue at the wound sites. Plants with the *Noble* LTR reporter construct showed more definite wound-specific GUS expression patterns.

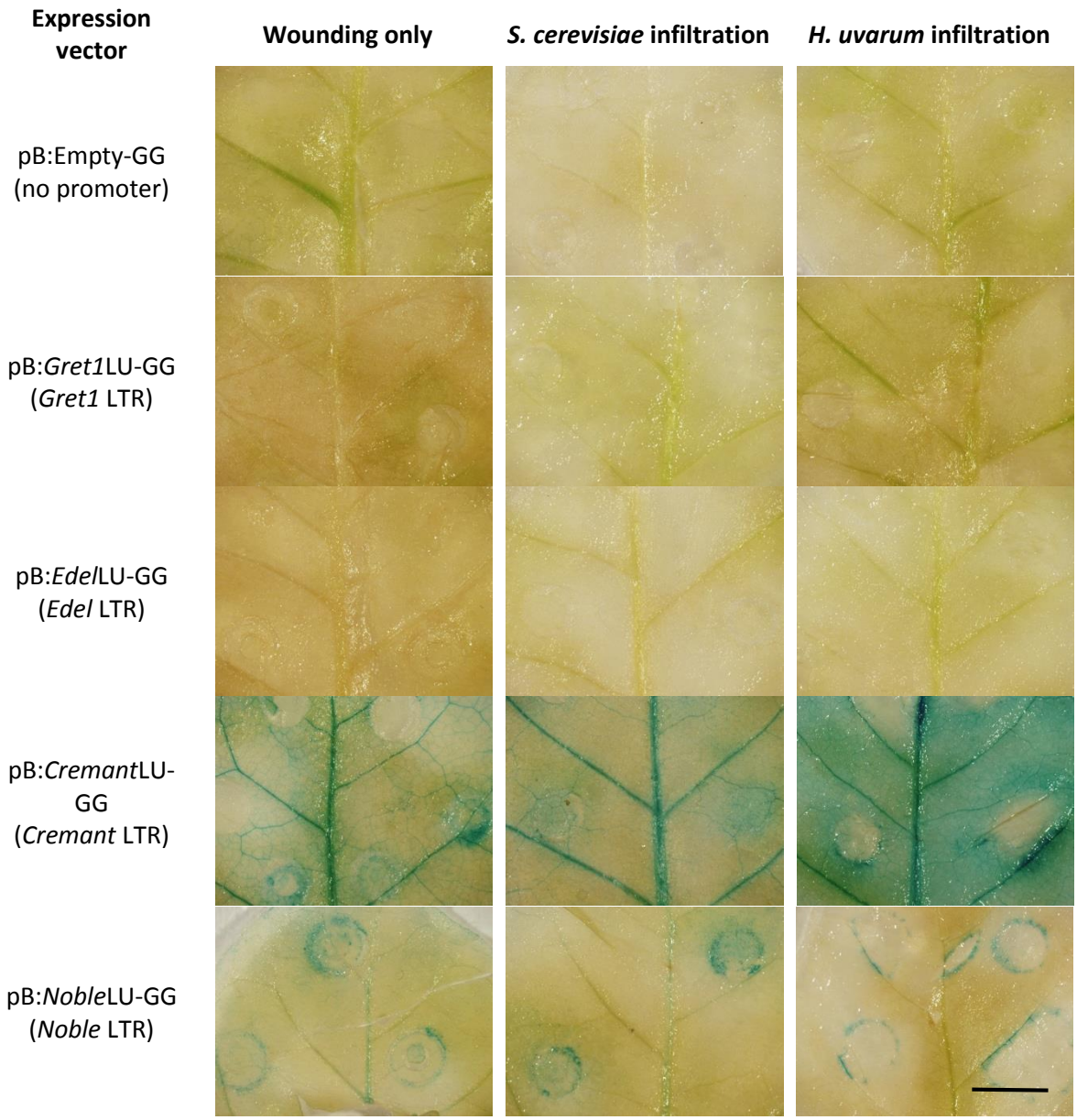


Figure 6.15 Expression patterns from retrotransposon LTRs in transgenic tobacco. Stable tobacco lines carrying *V. vinifera* retrotransposon LTR+UTR sequences preceding a GUS/GFP reporter gene were tested for GUS expression following stress. Leaves were either wounded with a syringe, infiltrated with live *S. cerevisiae*, or infiltrated with live *H. uvarum*. Each image is representative of three replicate experiments. Scale bar = 5mm.

LTR-reporter expression in grapevine leaves

As with the experiments that used stable transgenic tobacco plants, the addition of yeast cells to *A. tumefaciens* cultures during transient transformation of grapevine had no effect on the activity of retrotransposon LTRs as promoters. In the case of grapevine, only the Edel LTR and the 35S promoter induced expression of the GUS gene in leaf tissue (Figure 6.16).

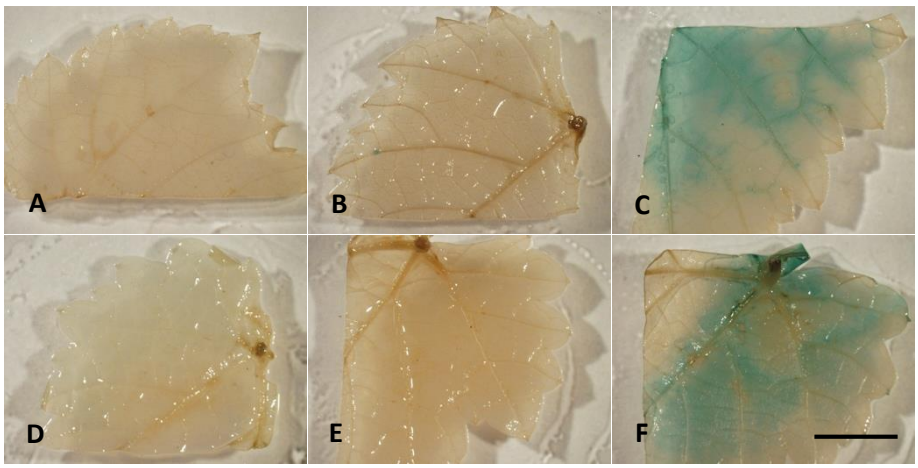


Figure 6.16 Transient transformation of young S. blanc leaves with LTR-GUS reporter constructs. The addition of *S. cerevisiae* or *H. uvarum* cells to the bacterial suspensions had no effect on GUS expression, therefore samples infiltrated only with *Agrobacterium* strains carrying binary expression vectors are shown. Key to transformation vectors: A: pB:Empty-GG, B: pB:*Gret1*LU-GG, C: pB:*Edel*LU-GG, D: pB:*Cremant*LU-GG, E: pB:*Noble*LU-GG, F: pB:35S-GG. Images are representative of 5 replicates. Scale bar = 5mm.

6.5 Discussion

6.5.1 Sample processing

Some degradation was seen in the RNA samples treated with live yeast but not in any of the control samples. The RNA integrity was unaffected by soaking tissue in an RNase inhibitor prior to extraction, indicating that the RNA degradation had not occurred during the extraction protocol. Rather, it is likely that a proportion of the cells had not survived the stress treatment, and as a result intact RNA could not be extracted from these.

Several reports have shown that the effect of moderate RNA degradation on transcript quantification is minimal to none when using small amplification targets and normalising to an internal reference gene (SCHOOR *et al.* 2003; FLEIGE AND PFAFFL 2006; RAVO *et al.* 2008). In the case of the experiments reported here, retrotransposon transcript abundance was normalised to two reference genes. Analysis of the relative transcript abundance of the reference genes with respect to one another revealed little variation across all samples.

A combination of random hexamers and oligo(dT) were used to prime the reverse transcription reactions in order to enable amplification of all retrotransposon-derived transcripts. However, qPCR reactions performed on cDNA produced from the same samples using only oligo(dT) to prime the reverse transcriptase reaction showed no significant difference in relative transcript abundance for any of the primer pairs (data not shown).

6.5.2 Wild yeasts stimulate retrotransposon expression

Due to the number of samples produced, it was not possible to quantify the response of each retrotransposon family to all stresses. Instead, the transcriptional responses of selected retrotransposon families were measured across the maximum range of stress types.

Several plant retrotransposon families are known to demonstrate significantly increased activity following abiotic and biotic stresses of the host tissue (see 2.2.5). In this study transcript levels of the four retrotransposon families analysed did not significantly increase following any of the abiotic stresses tested. In contrast, transcription of all four retrotransposon families increased markedly after co-cultivation of EC with yeast species collected from New Zealand vineyards. Although the level of transcription varied greatly among biological replicates, the same effect was observed in three separate experiments. In the first experiment (Figure 6.7), for example, *Gret1* transcription increased between 30 and 315-fold following cultivation with *S. cerevisiae* (yeast 1). These variations in the magnitude of the response could possibly be a result of physiological inconsistencies among the proliferating EC masses, as seen by the variations in appearance (Figure 6.9).

Surprisingly, a rise in retrotransposon transcription was not seen when the same tissues were exposed to either live *A. tumefaciens*, an extract of *Botrytis cinerea* or a *Trichoderma viride* extract that has been shown to increase transcription of the endogenous *Tnt1* retrotransposon in tobacco (Onozuka R-10). Although those treatments did not involve the use of live cultures, it was initially expected that extracts of pathogenic microorganisms would produce a greater shock to the tissue than live organisms that are not commonly associated with disease in grapevine. It is possible that neither of the fungal extracts is capable of applying sufficient stress to trigger the release of retrotransposition silencing. However, these extracts did appear to have an impact on cell health and viability, as seen by a slight browning of the tissue two days after treatment. Instead, it would appear that the increased retrotransposon transcription is not simply a general biotic stress response but rather one that was specifically triggered by the co-cultivation of EC tissue with yeast.

The mechanism by which retrotransposons may be transcriptionally responsive to yeast cultures is unknown. However, yeast cell wall extract has been shown to trigger pathogenesis-related (PR) genes in Arabidopsis (MINAMI *et al.* 2011). As discussed, retrotransposon sequences, including those analysed in this study, harbour multiple CREs common to PR genes and have frequently been seen to show a pathogen-responsive pattern of activity (see 2.2.5 and 3.4.3). While yeast are not generally considered pathogens of grapevine, there is evidence that wild yeasts may become pathogenic under certain circumstances (GOGNIES *et al.* 2001).

6.5.3 The contribution of native yeasts to *terroir*

The stimulation of transposon expression in grapevine by the yeasts tested is directly relevant to the goal of this project, which is to increase rates of transposon mutagenesis. Yet, the implications of this relationship between wild yeasts and grapevine have wider importance. It is well known that the sensory profiles of wines are affected by the location in which the grapes were grown. This origin effect, broadly described as *terroir*, is generally attributed to variation in climate and soil, in addition to the grape variety grown (VAN LEEUWEN *et al.* 2004; VAN LEEUWEN AND SEGUIN 2006). However, surveys of microbial diversity show that native yeast strains differ greatly between wine growing regions. The recognised contribution of these wild yeast strains to wine fermentation has led several researchers to argue that native yeast populations should be considered an aspect of *terroir* (VEZINHET *et al.* 1992; RASPOR *et al.* 2006; RENOUF *et al.* 2006; FLEET 2008).

The results of this study show that endogenous yeast cultures have the potential to influence the genetic and epigenetic characteristics of vines via the mobilisation of transposable elements. Such effects could be considered “neo-Lamarckian” in the sense that an environmental stimulus may result in heritable changes to the grapevine genome. It is important to note, however, that the transposon activation reported here was limited to embryogenic callus, a tissue type that does not

exist naturally in vineyards. Leaf tissue did not show the same response to yeast exposure. In order to determine whether native yeast species contribute to genetic or epigenetic changes in cultivated vines, it would be necessary to test whether they stimulate the transcription of TEs in other vine tissue and whether increased TE transcription ultimately results in new TE insertions. Of particular interest are those tissues that are more similar to EC in their pluripotent nature and from which new vegetative material originates, such as apical meristem. This is currently an area of interest to our research group.

6.5.4 General increase of multiple TE families

Despite the variation in the magnitude of the response between the TE families, samples that showed a response generally had raised transcript levels for all of the TE families tested. Transcript expression begins with the binding of transcription factor proteins and the assembly of the initiation complex, but this is usually suppressed in the case of transposons by epigenetic modifications (see 2.2.4). In order for TE mRNA to accumulate, host silencing must be released. Transposon activation resulting from a general hypomethylation of the DNA has previously been seen under two conditions. The first is during periods of environmental stress, when the risk of detrimental mutations is weighed against the necessity for adaptation (see 2.2.5). The second is in the vegetative tissues produced during meiosis, where TE activation is believed to contribute to silencing in the gametes and zygotic tissue (CHOI *et al.* 2002; HSIEH *et al.* 2009; SLOTKIN *et al.* 2009).

The multi-family enhancement of retrotransposon transcription suggests a state of less intensive retrotransposon silencing in the solid EC tissue. The effect was specific to both the EC tissue grown on solid medium and the yeast stresses, within the scope of these experiments. Evidence of such epigenetic changes leading to TE activation has previously been reported. Dedifferentiated cultures of Arabidopsis cells displayed hypomethylation of chromatin over time, together with an increased transcription of TEs (TANURDZIC *et al.* 2008). More recently, grapevine material has been seen to accumulate epigenetic changes when passaged through tissue culture (BARÁNEK *et al.* 2010). It is likely that the cultures used in this experiment have undergone the same DNA hypomethylation, with a consequent increase in retrotransposon transcription. Further experiments to determine the methylation status of transposons in EC tissue are needed to confirm this.

6.5.5 Difference between relative and absolute expression

Comparative analyses of transcript data show that the relative expression change was highest for the *Cremant* family of retrotransposons following yeast exposure (Figure 6.13). However, when absolute transcript abundances are compared (Figure 6.12), it is clear that *Cremant* was generally absent from the transcriptome in control samples, and increased to 0.5-3.6% of the mean abundance of the

reference genes in EC exposed to yeast cultures. *Noble*, in contrast, was present at 25-30% of the abundance of the reference genes in control samples, but increased to levels similar to Actin and EF1α following yeast stress (34-312% of the reference gene mean).

It could be that the variation in the degree of transcriptional response between the retrotransposon families is due to the different *cis*-regulatory elements in the LTR promoter sequences. However, it is not possible to confirm this from the available data, as the LTR promoters did not display the same yeast-induced response in the N. tabaccum plants. Also, no difference was observed with regards to the stress type responsible for inducing expression among the retrotransposon families tested. Samples in which a response was seen showed increased transcription of all TE families tested.

An alternative explanation is that the transcript number is related to the number of source elements that can become transcriptionally active following a general release of retrotransposon silencing. This hypothesis is supported by the fact that the absolute amount by which retrotransposon transcript numbers increased in the responsive samples corresponds with the abundance of complete elements for each TE family in the grapevine genome. For example, Cremant consistently showed the lowest absolute increase in TE transcript of the four families (Figure 6.12), and exists only as 8 complete insertions in the grapevine reference genome (Figure 3.4). Noble, which consistently showed the greatest absolute increase, is represented by 104 complete genomic copies. Due to the positioning of the qPCR primers (at the start of the 3'LTR; Figure 6.1), only transcripts originating from elements that are largely intact would have been detected in these experiments.

6.5.6 LTR-reporter transformations

Grapevine retrotransposon LTR sequences were effective in promoting reporter gene expression following transient transformation of the model organism *N. tabaccum*, though reporter genes preceded by the *Gret1* LTR showed the least expression. Among stable transgenic plants, only plants with the *Cremant* and *Noble* LTR constructs showed reporter gene expression. Yeast infiltration sites showed a higher degree of expression, but simply pressing a syringe against the leaf was enough to trigger increase expression of the reporter construct. This indicates that these LTR sequences are able to initiate transcription in response to stress in a heterologous system, although the specific yeast response could not be replicated in tobacco.

Though the presence of the transgene was confirmed by PCR in all cases, no expression was seen from the *Gret1* and *Edel* LTR sequences. It is likely that these transgene constructs became silenced within the tobacco genome. The phenomenon of transgene methylation silencing is well documented, particularly with regards to TE-associated sequences (MATZKE *et al.* 2000; BAULCOMBE 2004; FRANCIS AND SPIKER 2005; MEYER 2011).

Grapevine leaves transiently transformed with LTR-reporter constructs showed no expression except for in the case of *Edel*. It is worth noting that the *Edel* retrotransposon family has the highest number of complete elements in the grapevine genome. It appears that transcript expression from the *Edel* LTR sequence has not been as efficiently silenced within the grapevine genome as the other three retrotransposon families tested, which may explain the high number of recent genomic insertions. This could be due to the domestication of an *Edel* LTR as a functional promoter element in the grapevine genome.

The *Cremant* LTR sequence proved to be the strongest promoter within the heterologous tobacco system, but showed the least background transcript abundance in grapevine EC cultures and is represented by the fewest insertions in the grapevine genome. In contrast to the situation with *Edel*, this element family may be highly aggressive in its expression, resulting in a more rapid and thorough silencing within the grapevine genome.

The *Noble* LTR was also able to escape silencing in stable tobacco transformants. Interestingly, *Noble* was seen to have the highest background level of expression in grapevine EC cultures (Figure 6.12) and is the most abundant of the four retrotransposon classes in grapevine EST databases (see 3.4.4). This lack of silencing in two species may indicate that the LTR sequence of *Noble* elements is not easily recognised and targeted by *de novo* methylation systems within host cells. Consequently, *Noble* has recently been able to proliferate successfully in grapevine, and exists as the second most abundant family of complete elements (Figure 3.4).

6.6 Conclusions

The transcript abundance of each of the four retrotransposon families analysed increased greatly following exposure of grapevine cells to live yeast cultures. The response was only seen in EC tissue grown on solid medium, and varied greatly in amplitude between biological replicates. Surprisingly, bacterial cultures and extracts from pathogenic fungi did not induce the same response. The transcriptional activation of mobile elements by yeast strains endemic to vineyards indicates a potential neo-Lamarckian link between the vineyard environment and the hereditary material of the vines within the vineyard, however it remains to be demonstrated that transposition rates increase with elevated TE transcription and that this same effect occurs in adult grapevine tissues.

The four retrotransposon families selected for study in this chapter showed a concomitant response to the yeast treatment. Also, the absolute increase in transcript number reflected the number of complete elements present in the grapevine genome for each family. This is possibly indicative of a reduced silencing of the mobilome as a whole. The same parallel stimulation was not seen in the heterologous tobacco system, where *Gret1* and *Edel* LTRs were silenced in stable transgenic plants but *Cremant* and *Noble* showed a low level of activity that was increased at wound sites.

Transcriptional activation of grapevine retrotransposons in totipotent cells in response to exposure to native yeast species demonstrates the first step towards exploiting endogenous TE activity to generate somaclonal mutant vines. Vines will need to be regenerated from this tissue in order to detect whether complete transposition has taken place and, if so, the effect that this has on the genotype and phenotype of the plants.

Chapter 7

The Generation of a Tagged Mutant Population of Pinot Noir Vines

7.1 Overview

Four retrotransposon families have been shown to be active in embryogenic callus (EC) cultures of Pinot noir clone UCD5 following stress treatment of the tissue (Chapter 6). Although transcription of transposable elements (TEs) may lead to epimutations (DNA methylation and histone modifications), change to the genetic code requires reverse transcription and insertion of these elements at new loci. In order to test whether complete transposition had occurred, 160 vines regenerated from EC cultures were pooled into groups of five samples and screened by S-SAP (see Chapter 4). Using five primer combinations to amplify genomic fragments flanking each of four retrotransposon families (*Gret1*, *Edel*, *Cremant* and *Noble*), 24 new unique polymorphisms were detected.

Fifteen unique phenotypes were recorded among the regenerated population of vines, although five of these disappeared as the plants aged. Further phenotypic variation is expected to become evident in later growth seasons, as maturation of the vines allows post-juvenile traits such as flowering and berry development to be scored. Retrotransposon mutations have previously been identified as the cause of specific grapevine bud sport mutations and have been shown to account for the majority of inter-varietal phenotypic variation. However, this population provides the first example of real-time mutagenesis in grapevine by the stimulation of endogenous transposable elements to produce new genetic diversity.

Further characterisation of the loci affected by transposition events and the total level of genomic variation among the regenerated plants requires whole genome sequence comparison.

7.2 Introduction

7.2.1 Tagged mutant populations

As discussed (see 2.1.5), the production of mutant plant populations by a variety of methods has been effective in generating novel phenotypes of many species. While approaches such as EMS and radiation mutagenesis are effective in producing random mutations, identification of individual mutations against the background of an entire genome is not a trivial task. Tagged mutagenesis overcomes this obstacle by producing mutations that are linked to specific nucleotide sequences. A common example of this is the mutagenesis of plant genomes by the insertion of T-DNA during Agrobacterium infection (JEON *et al.* 2000; ALONSO *et al.* 2003).

Since the work of Barbara McClintock (1948) the mutagenic activity of TEs has been used for functional genomics studies. The small sizes of bacterial genomes have made them a particularly attractive target for random mutagenesis. Consequently, dozens of bacterial genomes have been screened for virulence associated genes using transposon mutagenesis (HENSEL *et al.* 1995; SAENZ AND DEHIO 2005). In plants, large-scale transposon saturation mutagenesis experiments have been undertaken in several species using both endogenous and transgenic TEs (see 2.2.8).

Although the aim of genome saturation mutagenesis experiments is generally to generate representative mutations in each gene, complete saturation of the genome is an elusive goal in a finite population. As the saturation approaches 100%, the populations sizes required to increase the saturation further grow exponentially. For this reason, 'saturation' is generally defined as a high probability (e.g. 95%) of finding an individual in the population with a mutation at any given locus.

7.2.2 Transposon mutagenesis approaches

The first transposon mutagenised populations were produced by using plant lines in which a particular TE family is unusually active, as in the case of McClintock's maize lines (MCCLINTOCK 1951). However, this approach relies on the identification of a genotype with highly active mobile elements that can be crossed with the target genotype. Once mutations have occurred, further crosses are required to remove the genes necessary for transposition, in order to stabilise the genotype. This is particularly important in the case of class II TEs, which can leave their insertion site at a later stage, resulting in a phenotype reversion. The suitability of such approaches, which require multiple rounds of crossing, is limited in long-lived perennial crops, such as grapevine, by long generation times and high genotypic heterozygosity.

More recently, researchers have used transgenic techniques to move active TEs across species boundaries, in order to further exploit their mutagenic capacity. This approach has been successfully

used to create tagged mutant populations in many species (see 2.2.8). The primary goal of these collections has been to define phenotypes associated with mutations at each gene. While the value of mutagenesis to functional genomic studies is well established, the artificial introduction of foreign genetic material constitutes the creation of genetically modified (GM) plants. Such species are the subject of strict regulatory controls and controversial public opinion in most countries, which limits the industry implementation of lines displaying desirable traits.

However, the genomes of all eukaryotes studied to date show evidence of historical endogenous transposon activity. Within grapevines, multiple TE families exist which appear to have been recently active and retain the functional sequences required for transposition. Given that existing endogenous TEs appear to be a major source of natural genetic variation, transposon mutagenesis should be possible without the need to introduce foreign TEs by sexual crossing or transgenesis.

7.2.3 TE activity in response to stress

To limit the detrimental effects of their mutagenic nature, endogenous TEs are usually strictly silenced within host genomes by DNA methylation and histone modifications (see 2.2.4). However, this silencing is not permanent, and TE activity can be released in times of stress or in particular tissue types (see 2.2.5). While stress-responsive stimulation of transposition in grapevine has not yet been reported, methylation patterns of grapevine DNA have been shown to be altered (with an overall decrease in methylation) by both tissue culture and thermotherapy at 37°C (BARÁNEK *et al.* 2010).

7.2.4 Somatic embryos as explant material

Mutations that occur in grapevine vegetative material generally only affect a single cell layer, resulting in periclinal chimeric plant tissues (FERNANDEZ *et al.* 2006). In order to minimise the risk of chimerism among mutant plants, embryogenic callus (EC) cultures were used as a target material for retrotransposon stimulation. It has been shown that vines regenerated from EC cultures do not maintain the chimerism of the parent tissue used to initiate those cultures (FRANKS *et al.* 2002; BERTSCH *et al.* 2005). Furthermore, the results given in Chapter 6 indicate that four retrotransposon families show a transcriptional response to stress in these cells that was not seen in vegetative tissue.

Retrotransposon-based marker systems have proved effective for detecting genetic variation in clonal sports of clementine (BRETÓ *et al.* 2001) and apple (BRETÓ *et al.* 2001; VENTURI *et al.* 2006). This chapter describes the regeneration of a population of Pinot noir vines from cell cultures exposed to stress events shown previously to increase retrotransposon transcription (see Chapter 6). Insertion

profiles generated for four retrotransposon families according to the protocol described in Chapter 4 show novel genetic variation within the regenerated plants.

7.3 Methods

(For media recipes see Appendix D)

7.3.1 Embryogenic callus initiation

Several described methods (reviewed in 2.3.2) for the initiation of grapevine embryogenic callus were tested. One hundred leaf disks from young leaves of both Pinot noir UCD5 and Chardonnay Mendoza plantlets were cultured according to the protocol of Das and colleagues (2002). Likewise, one hundred sterilised whole flowers of P. noir UCD5 and Chardonnay Mendoza were cultured according to the protocol of Gambino and co-workers (2007).

The initiation and maintenance of embryogenic callus (EC) culture from the anthers of unopened flowers has already been described (see 6.3.2). The protocol used involves plating excised anthers on B medium for EC initiation (PERRIN *et al.* 2004), followed by long-term calli maintenance and proliferation on C_1^P medium (TORREGROSA 1998). Embryogenic callus cultures of four clones (P. noir UCD5, P. noir AM10/5, S. blanc UCD1 and Chardonnay Mendoza) were initiated in this way.

7.3.2 Shock treatments and vine regeneration

Embryogenic callus cultures of Pinot noir UCD5 were selected for environmental shock treatments and regeneration of new plants. Shock treatments were performed as described previously (see 6.3.3). For each shock treatment 0.5g of EC cells were exposed to shock conditions for 48hr on C_1^P medium. Cells were then directly transferred to solid EG medium plates for germination under standard growth conditions (25°C, 16hr:8hr light:dark cycle) with monthly transfers to fresh medium. When distinct individual embryos became visible (late heart-stage) and began to show green colouration, these were separated from the developing masses and cultured on EG medium individually.

Germinating plantlets approximately 10mm in length were transferred to sealed tissue culture tubs (6cm high x 8cm diameter) containing 50ml EG medium and incubated under the same growth conditions. When plantlets were approximately 5cm in height, they were removed from the medium and potted in Black Magic potting soil (The Scotts Company LLC) in 7cm x 7cm x 7.5cm (l x w x h) pots. The potted plantlets were incubated in sealed 35cm x 20cm x 18cm (l x w x h) tissue culture chambers for 2 weeks under standard growth conditions, after which the chamber lids were removed. When vines had grown to a height of approximately 20cm, they were transferred to 2.5L pots containing long-term potting mix (Appendix D) and moved to a 25°C shadehouse.

7.3.3 S-SAP on pooled DNA samples

The dilution factor at which a particular polymorphism can be detected within a pooled DNA sample was determined by combining young leaf tissue of P. noir UCD5 and Chardonnay Mendoza vines at varying ratios by weight (1:0, 1:4, 1:9, 1:14 and 0:1) and grinding to a powder in liquid nitrogen. A column-based extraction kit (NucleoSpin Plant II kit; Machery-Nagel) was used to extract pooled DNA from 100mg of this tissue.

For retrotransposon insertion profiling of vines regenerated from EC, young leaf tissue (approximately 3cm diameter leaves) of five plants was combined in equal amounts by weight and ground together in liquid nitrogen. Genomic DNA was extracted from 100mg of each of the pooled sample tissue. The number of DNA samples prepared (each representing five plants) is given in Table 7.1.

Table 7.1 Total numbers of Pinot noir vines produced and the number of DNA samples collected from these. One sample represents 5 pooled vines.

Treatment	Regenerated vines	Pooled DNA samples
Control	23	4
50mM salt	10	2
100mM salt	11	2
4°C	17	3
37°C	17	3
Y1: *Saccharomyces cerevisiae*	20	3
Y2: *Pichia kluveri*	17	3
Y3: *Aureobasidium pullulans*	18	3
Y4: *Hanseniapora uvarum*	15	3
Y5: *Rhodotorula glutinis*	15	3
Y6: *Cryptococcus magnus*	20	3
TOTAL	*183*	*32*

Genotyping was performed using the optimised S-SAP protocol described (see Chapter 4). Briefly, 500ng of DNA was co-digested with 5U of *Mse*I and *Eco*RI for 3 hours. Double-stranded oligonucleotide adapters were ligated to the digest products and adapter-specific primers were used for preamplification. Diluted preamplification products were used for selective amplification with fluorescently tagged LTR primers and *Mse*I adapter primers with two selective bases. Products from the selective amplification of four retrotransposon classes were multiplexed for sizing by capillary electrophoresis.

For each DNA sample, twenty selective amplification reactions were performed. Each reaction used a different combination of one LTR-specific primer and one *Mse*I adapter-specific primer with two selective bases. Amplifications were performed using the adapter primers Mse(TC), Mse(TA) and Mse(TG), paired with the labelled primers amplifying downstream of retrotransposon LTR sequences (Gret1_DS, Edel_DS, Cremant_DS and Noble_DS), as well as Mse(TT) and Mse(TG) primers paired with labelled primers amplifying upstream of LTRs (Gret1 rev(0), Edel rev(0), Cremant rev(0) and Noble rev(0)). All primer sequences are given in Appendix C).

7.4 Results

7.4.1 Initiation of embryogenic callus

Cultured disks of Pinot noir and Sauvignon blanc leaved produced a light-brown, wet callus during three weeks of cultivation. After transferring the callus from low light to high light conditions, compact white masses appeared within the calli. These masses did not develop any further within the calli, and did not survive when removed.

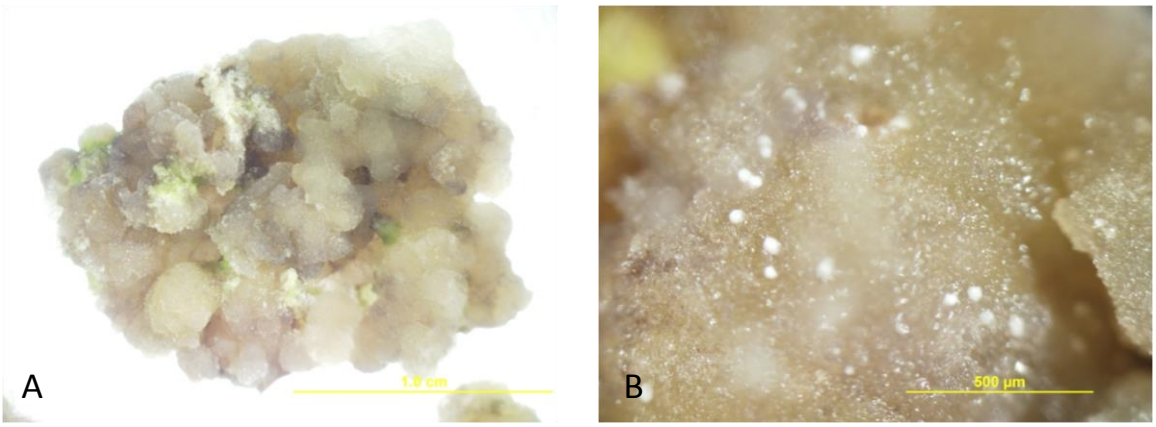


Figure 7.1 Embryogenic callus culture from leaf disks A: Pinot noir disks callus after 1 month (DAS *et al.* 2002) B: Dense, white masses within non-embryogenic callus.

Whole flowers and excised ovaries produces a brown, loose callus after approximately one month of tissue culture, but no EC was observed from these tissues. In contrast, white, compact EC developed at the cut site of excised anthers after 1-2 months (Figure 7.2). Although the efficiency of EC initiation was below 10% for all genotypes (Table 7.2), EC masses proliferated rapidly after transfer to C1P medium.

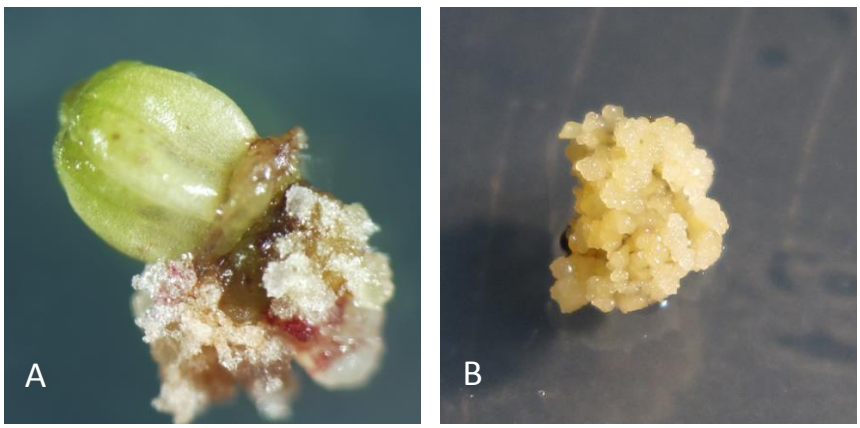


Figure 7.2 Non-embryogenic callus (A) derived from whole flowers and embryogenic callus (B) derived from cultured anthers of Pinot noir UCD5.

Table 7.2 Efficiency of embryogenic callus induction from anther cultures of four grapevine clones.

Grapevine clone	Anthers plated	Number of EC	Initiation efficiency
P. noir UCD5	2477	72	2.90%
P. noir AM 10/5	360	5	1.39%
S. blanc UCD1	896	55	6.14%
Chardonnay Mendoza	1229	46	3.74%

7.4.2 Post-shock regeneration of somatic embryos

Germination of somatic embryos began within two weeks after transferring EC to hormone-free medium for all of the varieties tested (Pinot noir, Sauvignon blanc and Chardonnay; Figure 7.3). Adding activated charcoal (3g/L) to the media and maintaining time between transfers to less than 1 month minimised browning of tissue. No significant effect on germination timing or efficiency was observed when BAP was added to the regeneration medium, as recommended by Iocco and colleagues (2001).

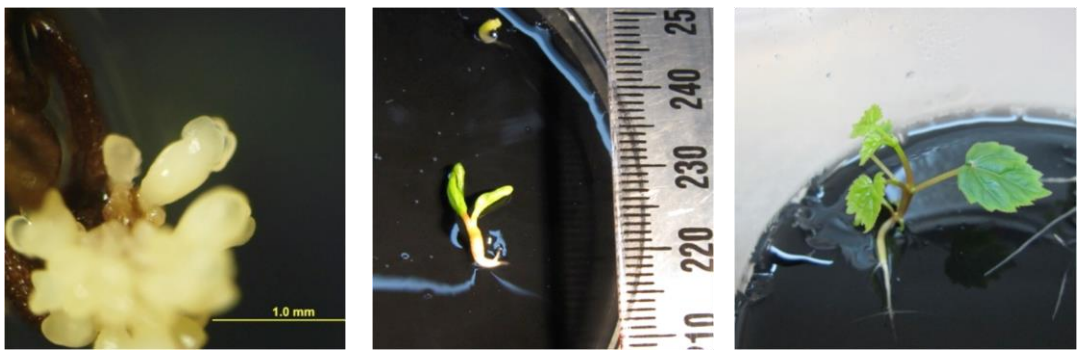


Figure 7.3 Somatic embryos of Pinot noir from control treatments germinating after transfer to hormone-free medium.

The exposure of P. noir EC to certain shock treatments significantly affected the germination and growth of plantlets from these tissues. No plants were recovered from EC shocked with 200mM NaCl, UV-C light or *A. tumefaciens*. Cells treated with Onozuka R-10, 100mM NaCl and *Botrytis* extract showed severely hindered germination and growth (Figure 7.4). Cells treated with 4°C shock began germination immediately after transfer to EG medium, and development of these plantlets remained approximately 1-2 weeks ahead of the control samples.

Control | *Botrytis* extract | 100mM NaCl

Figure 7.4 Regenerating Pinot noir vines displaying morphological and developmental differences eight weeks after stress exposure. Plantlets treated with *Botrytis* extract and 100mM NaCl shock show delayed development and colouration.

7.4.3 Phenotypic variation of regenerated vines

In addition to alterations to the germination period, plantlets regenerated from stressed embryogenic callus frequently showed phenotypic abnormalities not seen in the untreated samples (Figure 7.5). Sixteen plants were characterised as phenotypically distinct from the typical Pinot noir UCD5 type by visual analysis. Of these, five variant phenotypes disappeared as the plants matured and six were only detected in the mature plants. The others appeared early in the development of the plants and remained stable in the potted vines. A catalogue of the mutant phenotypes is given in Table 7.3.

Figure 7.5 Development of a Pinot noir grapevine plantlet displaying an abnormal leaf shape phenotype. The first eight leaves produced had irregular shapes and reduced surface area. After three months of growth, all further leaves showed normal morphology. This plant was regenerated from EC stressed by exposure to yeast 4 (*Hanseniapora uvarum*).

Table 7.3 Phenotypic variance among regenerated vines. The "plant" identifier name indicates the stress treatment to which the embryos used to regenerate the plant were exposed and the vine's number.

Plant	Description	Photographs
Control 1	Short & bushy	
Control 10	Control	
Control 12	Rounded leaves	
Control 17	Rounded, bumpy leaves	
Control 20	Bumpy leaves, grew out later	

50mM NaCl-7	Bud burst at every node	
37°C - 10	Bumpy, odd-shaped leaves	
37°C - 13	Strange leaf texture (possible lack of cuticle). Low mildew resistance	
Yeast1-9	Unusual leaf shape and texture	
Yeast2-7	Red venation, grew out	
Yeast2-10	Spikey leaf shape. Grew out later	

Yeast3-2	Deep leaf sinuses	
Yeast3-13	Deep leaf sinuses, not distinguish-able later	
Yeast4-6	Deformed leaf shape. Grew out later	
Yeast4-9	Irregular leaf shape	

7.4.4 Transposon insertion profiles

Pooled samples of P. noir UCD5 and Chardonnay Mendoza tissue showed that polymorphic S-SAP markers are clearly visible when present in one-fifth of the pooled tissue. Regenerated vines were therefore combined in pools of five to screen for new insertions.

The S-SAP profiles showed a vey low proportion of polymorphic bands (Figure 7.6). A total of 24 unique polymorphic bands (Table 7.4) were identified in the S-SAP profiles of 32 pooled samples, representing 160 plants. Of these, 5 (21%) were associated with the *Gret1* family of elements, 13 (54%) with *Edel*, 4 (17%) with *Cremant* and 2 (8%) with *Noble*.

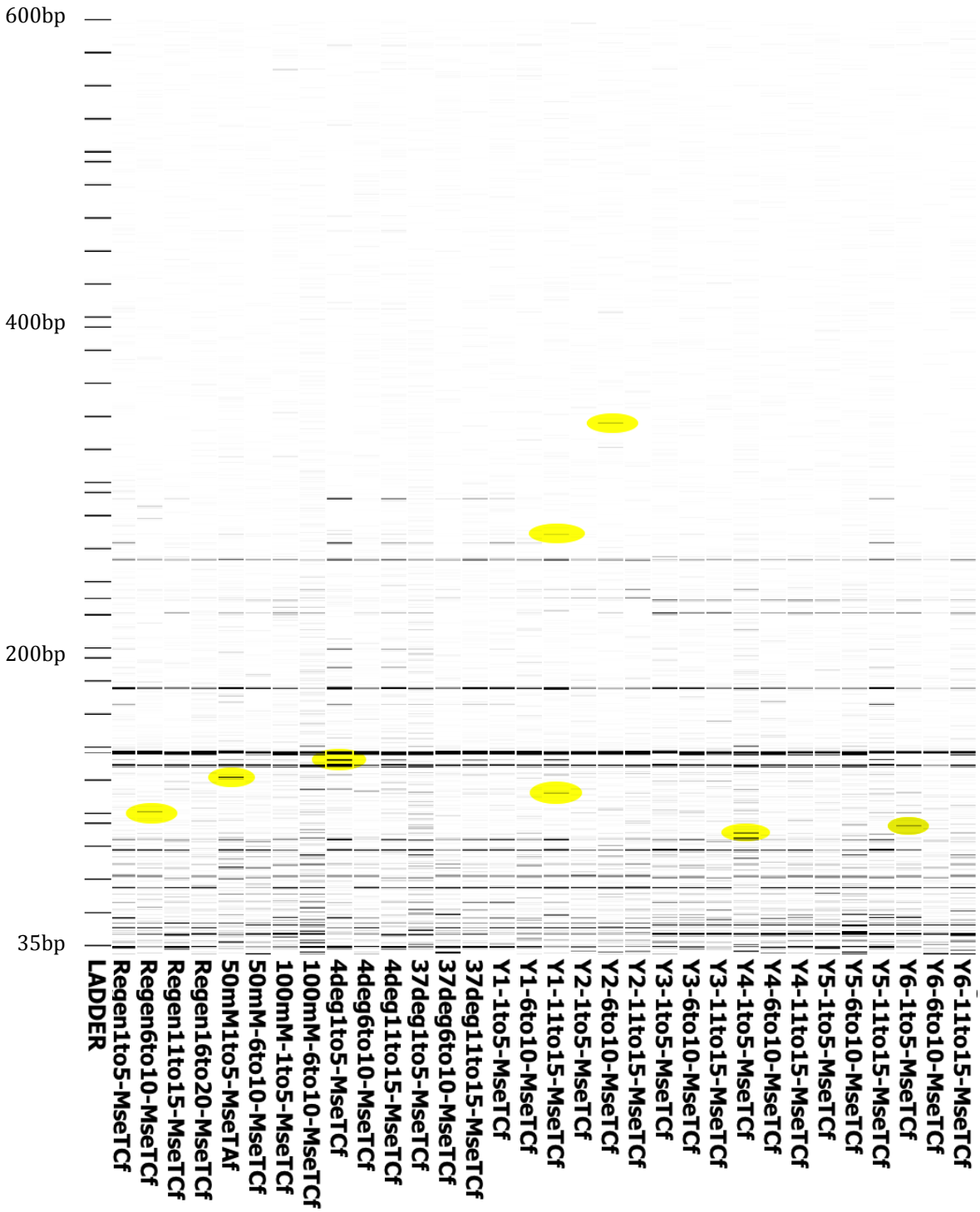
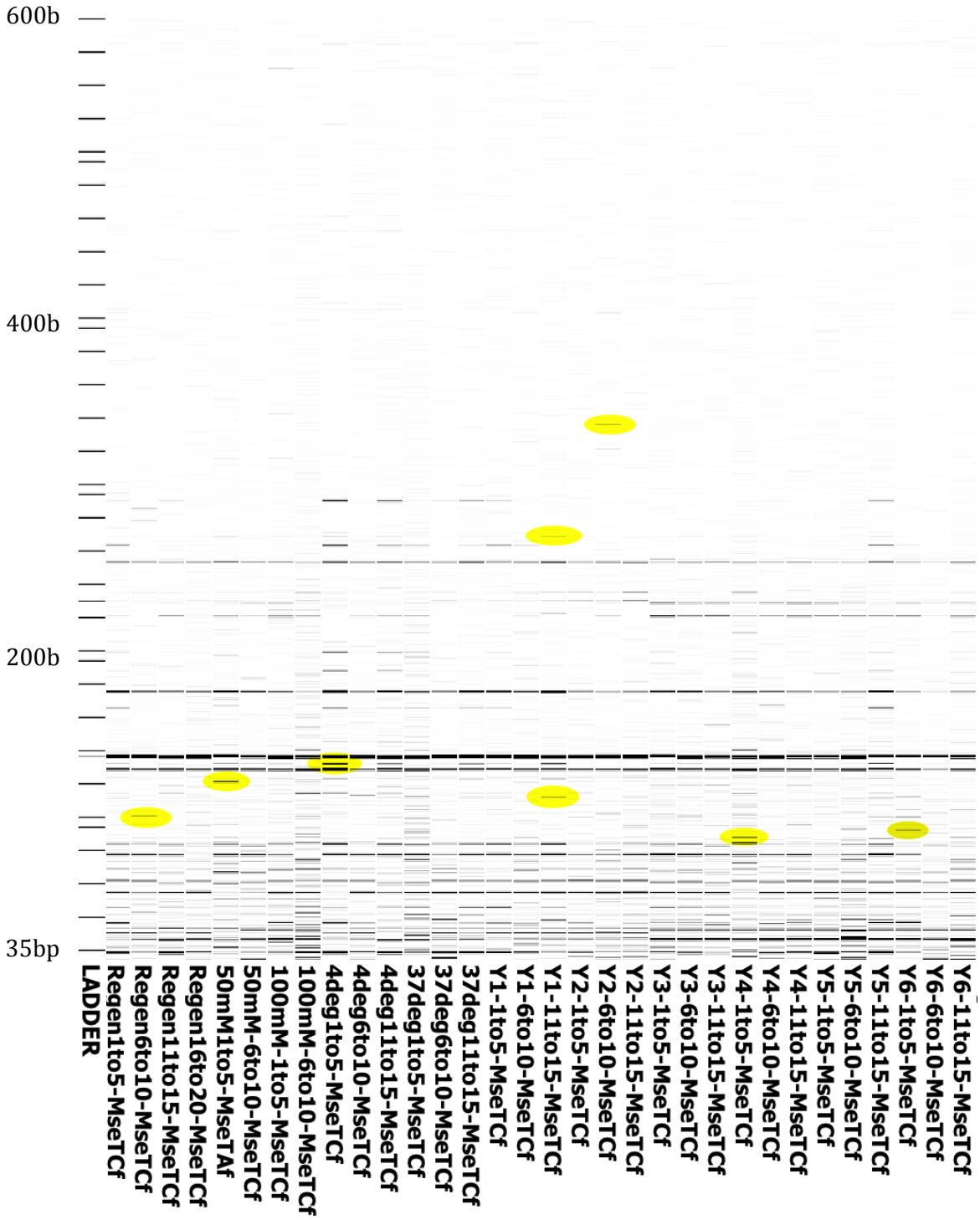
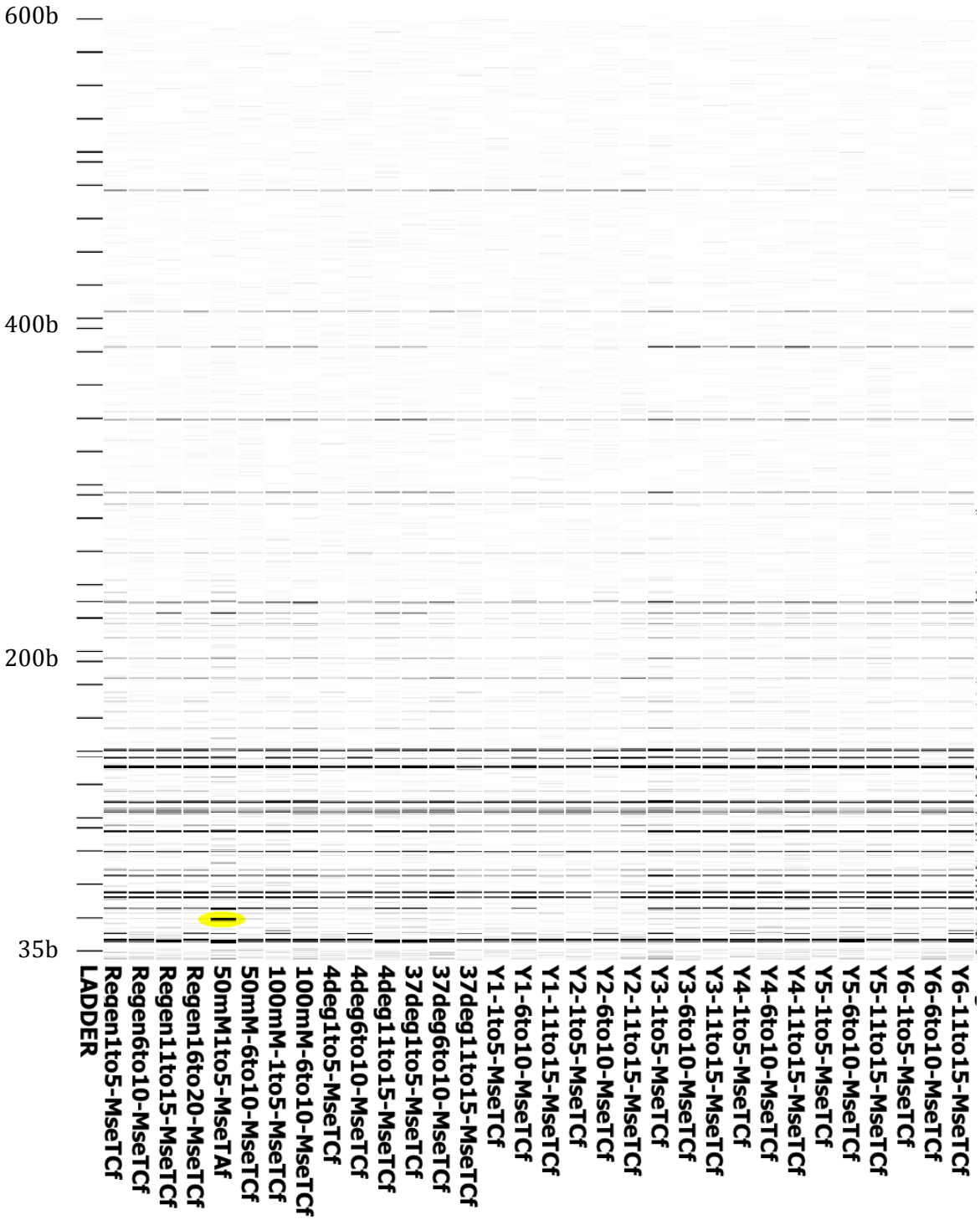


Figure 7.6 S-SAP profiles of 32 pools of DNA (five regenerated Pinot noir vines per pool).

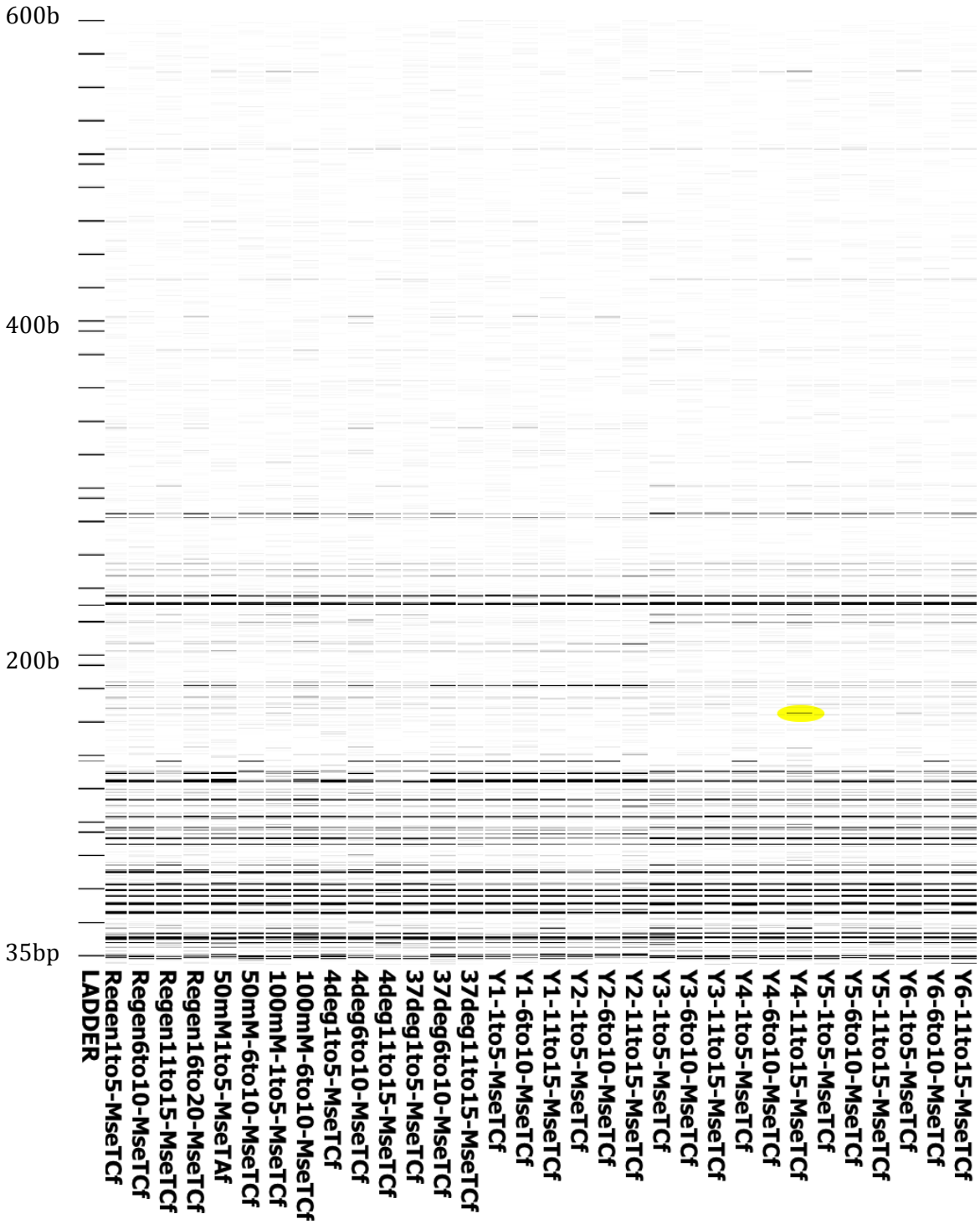
A) Gret1_DS & Mse(TC) primer combination. Each profile (lane) is produced by the amplification of genomic DNA from five pooled plants. Polymorphic insertions are highlighted in yellow. Key to lane names: Ladder: LIZ 1200 size standard, Regen: vines regenerated from unstressed EC, 50mM/100mM: vines regenerated from EC stressed with NaCl at 50mM/100mM concentration, 4degC/37degC = vines regenerated from EC stressed by incubation at 4°C/37°C, Y1-Y6: vines regenerated from EC stressed by exposure to *S. cerevisiae*, *P. kluveri*, *A. pullulans*, *H. uvarum*, *R. glutinis*, *C. magnus* respectively.



B) Edel_DS & Mse(TC) primer combination.



C) Cremant_DS & Mse(TC) primer combination.



D) Noble_DS & Mse(TC) primer combination.

Table 7.4 Polymorphisms identified within the population of regenerated P. noir vines. In each of five transposon display experiments all of the DNA samples were screened using the TE-specific primer and an adapter-specific primer with two selective bases. Each polymorphic DNA sample is comprised of the combined DNA of five plants regenerated from EC exposed to the same stress treatment. Only pooled samples displaying polymorphisms are listed.

Adapter primer	Transposon Primer	Polymorphic DNA sample		Fragment size (bp)
		Stress	Pooled plants	
Mse(TA)	Edel_DS	100mM NaCl	1-5	186
		4°C	1-5	154
		P. kluveri	6-10	356
		A. pullulans	11-15	199
	Noble_DS	*H. uvarum*	11-15	457
Mse(TC)	Gret1_DS	Control	16-20	145
		50mM NaCl	1-5	72
	Edel_DS	50mM NaCl	1-5	142
		S. cerevisiae	11-15	132, 289
		P. kluveri	6-10	356
		H. uvarum	1-5	108
		C. magnus	1-5	113
	Cremant_DS	50mM NaCl	1-5	60
	Noble_DS	*H. uvarum*	11-15	185
Mse(TG)	Gret1_DS	4°C	1-5	243
	Edel_DS	50mM NaCl	1-5	209
		4°C	1-5	116
		S. cerevisiae	11-15	108
		P. kluveri	1-5	152
	Cremant rev(0)	*P. kluveri*	6-10	242
Mse(TG)	Gret1 rev(0)	Control	16-20	65
		H. uvarum	1-5	65
Mse(TT)	Cremant rev(0)	*S. cerevisiae*	1-5	13
		A. pullulans	6-10	502

7.5 Discussion

7.5.1 Somatic embryos culture

Although Kikkert and colleagues reported similar EC initiation efficiencies between ovaries and anthers of Pinot noir clone 29 (2005), none of the ovaries cultured in the experiments described here were seen to produce EC. Leaf disks of Pinot noir and Sauvignon blanc, found to be effective for initiation of somatic embryos in four Indian grapevine varieties (DAS *et al.* 2002), did produce white masses with similar appearance to somatic embryos, but these masses did not grow further and could not be maintained.

Initiation of EC from the excised anthers of unopened flowers proved effective for all four clones tested. Although this technique does require dissection of unopened flowers in order to remove anthers, it is the most widely reported technique for initiation of EC cultures and was found to be most reliable for the varieties tested here. This explant tissue was also the only type from which EC were seen to grow in any of the genotypes tested in these experiments. With experience approximately 3 plates of 100 anthers could be prepared per hour. At the recorded efficiencies, EC could be initiated at an average rate of 1-6 masses per plate, depending on the genotype. Once initiated, callus lumps grew large enough to be divided into approximately 4-5 new lumps at each monthly subculturing. In this way a limited number of initial EC masses could be propagated so as to allow sufficient material for regeneration of several hundred plants within a matter of 6-12 months. Embryogenic calli maintained for approximately two years with monthly subculture on C_1^P medium showed no detectable changes to germination efficiency.

7.5.2 Phenotypic variation among regenerated vines

Several phenotypic traits that were particularly outstanding could be identified among the regenerated vines. These included stable abnormalities in leaf shape, leaf texture and internode distance. Due to disparities in germination and growth rates among regenerated plantlets, a four month period spanned the time between which the first and the last of the regenerated vines was moved from tissue culture conditions to the nursery. The different growth conditions rendered many other important phenotypic traits, such as vigour and stem diameter, impossible to compare during this first growth season.

Other important vine characteristics that can only be scored in future seasons include those pertaining to flowering and fruit development (e.g. number of bunches, date of ripening, berry chemical composition). Variations in these attributes are of particular significance to the yield and quality of the fruit at harvest. Changes to grape bunch architecture that result in increased air space

between berries would also be of particular interest, as these also likely to decrease susceptibility of fruit to bunch rot diseases, most commonly caused by *Botrytis cinerea*.

As mentioned (see 7.4.3), certain phenotypic abnormalities were initially distinct, but gradually disappeared over time. The abnormal leaf shape phenotype shown as an example of this (Figure 7.5) had completely disappeared after approximately three months of growth. One possible explanation for such phenotypes is that they are the result of genetic mutations later targeted for epigenetic silencing. The insertion of TEs has been shown on multiple occasions to increase the expression of nearby genes, either by introducing new promoter sequences (see 2.2.6), or by disrupting repressor elements (FERNANDEZ *et al.* 2010). However, transposon insertions can also attract host silencing mechanisms leading to the epigenetic silencing of adjacent genes (LIPPMAN *et al.* 2004; BIEMONT AND VIEIRA 2006). It is possible that as host silencing mechanisms target new insertion events, aberrant expression is stopped and phenotypic reversions occur. The eukaryotic silencing mechanisms that act within genomes are so effective at silencing TEs that this is now thought to have been their initial purpose (see 2.2.4).

7.5.3 Novel transposon mutations

Due to the high numbers of pre-existing TE insertions, S-SAP amplification reactions were performed using primers with two arbitrary nucleotide bases added to the 3' end. Each primer combination therefore displayed only $1/16^{th}$ of the total number of insertions for each retrotransposon family. However, within the data obtained from amplification of four retrotransposon families with five primer combinations, 24 novel transposition events were detected among the 160 plantlets screened. This amounts to an average of 0.48 mutations per genome from these four element families.

The actual rate of genetic variation is likely to be higher than that recorded among the regenerated vines, due to the negative genotoxic effects of insertions affecting critical genes. In addition to the plantlets lost due to lethal mutations, several plantlets that could be maintained in tissue culture did not survive hardening off, possibly indicating reduced fitness of the plants caused by new TE insertions.

7.5.4 The limitations of S-SAP

A thorough characterisation of genotypic change in the regenerated vines with respect to transposition events would require a comprehensive record of all new insertions. However, it is readily apparent that S-SAP is not ideally suited to this task.

The addition of selective nucleotides to the adapter primers reduces marker complexity, while limiting the subset of retrotransposon-flanking regions that are amplified. Nevertheless, certain insertions will not produce S-SAP fragments from any combination of selective bases. Typical S-SAP experiments begin with the co-digestion of genomic DNA with two restriction enzymes (*Eco*RI and *Mse*I in the case of the experiments reported here). For genomic loci in which an *Eco*RI restriction site exists between a retrotransposon and the nearest *Mse*I site, amplification using the combination of an *Mse*I adapter primer and the relevant retrotransposon LTR primer will not produce a product. Furthermore, if the distance between the retrotransposon primer and the nearest *MseI* restriction site is not within the range of fragments that can be properly size calibrated using the internal size standard (approximately 35bp to 1200bp in the case of the LIZ1200 standard), the amplified product cannot be compared between samples.

In spite of these limitations, S-SAP remains a useful molecular marker tool. The high inter-varietal polymorphism of S-SAP markers makes this approach a quick and cheap method for genotyping grapevine material based on variety. Furthermore, application of this method allowed easy detection of new retrotransposon polymorphism within a population of individuals. Being PCR-based, the protocol requires only standard molecular biology techniques and access to fragment analysis by capillary electrophoresis. Data analysis is relatively straightforward, as amplification products can be automatically size calibrated and directly compared in the form of pseudogels.

The sensitivity of fluorescent marker detection by a capillary sequencer allowed identification of unique heterozygous dominant markers in a pools of up to five diploid genomes (i.e. alleles which are present at a one-tenth concentration). However, recovery of insertion-site sequence data for S-SAP markers requires that fragments be separated by acrylamide gel electrophoresis (as described in 4.3.5). The reduced sensitivity of gel-based assays would reduce the number of samples that can be pooled for genotyping, thereby increasing the number of assays that need to be performed.

Nevertheless, the value of a mutant population depends on the ability to map novel genetic diversity to defined genomic loci. Furthermore, mutations caused by retrotransposon families other than those targeted by S-SAP will remain undetected unless an unbiased genotyping approach is adopted. A summary of the time and costs involved in performing S-SAP (Appendix B.6) shows that a complete analysis of only four TE families costs approximately NZ$230. The immense data output and falling price of shotgun DNA sequencing means that next-generation sequencing (NGS) techniques will likely prove to be not only faster and more comprehensive, but also cheaper than PCR-based assays for characterising TE polymorphisms mutant populations on a large scale.

7.6 Conclusions

Previous work (see Chapter 6) showed that four families of class I transposable elements are transcriptionally active in stressed EC cultures. In this chapter, retrotransposon insertion profile data demonstrated that genotypic variation associated with these same TE families is apparent in vines regenerated from the stressed cultures. This serves as the first reported evidence that endogenous grapevine retrotransposons can be stimulated to produce new genotypic variation.

Phenotypic variation was seen within the population of regenerated vines, however specific links between phenotypic change and transposition events have not yet been established. While molecular marker assays provide evidence of genotypic variation, mutant loci need to be mapped to their genomic context in order for preliminary predictions to be made regarding the possible effect of any new insertions.

Chapter 8

The Whole-Mobilome Analysis of Grapevines Regenerated from Embryogenic Cultures

8.1 Overview

The S-SAP analysis of grapevines regenerated from stress-treated embryogenic callus cultures indicated the presence of novel insertions for four retrotransposon families (Chapter 7). These families had been selected for analysis based on a computational search of the grapevine reference genome intended to identify recently active elements. To obtain a more comprehensive picture of transposon mobility within the regenerated vine population, whole genome resequencing was performed on twenty vines regenerated from somatic embryo cultures. The results show a broad mobilisation of transposable elements (TEs) from all superfamilies in the grapevine genome, except helitrons. An average of 64 new insertions per genome were observed, with class I and class II transposons accounting for 57% and 42% of these respectively. Plants regenerated from embryogenic callus (EC) stressed with extracts of *Pichia kluveri*, a yeast native to New Zealand vineyards, showed 63% more TE insertions than those regenerated from control EC cultures. These data are in agreement with previous results showing increased TE transcripts in EC co-cultivated with endemic yeast cultures (Chapter 6).

Fewer insertions were found within coding sequences than expected by chance. Only 2% of novel insertions were identified in these regions, which comprise 6% of the genome. A more moderate negative bias was seen against insertions in untranslated gene regions and introns. Those retrotransposon families previously identified as having most recently been active (Chapter 3) did not account for the highest number of new class I element insertions. This demonstrates the importance of unbiased analysis techniques when characterising mutations caused by endogenous transposons.

This is the first work to describe the plasticity of the grapevine mobilome as a whole. It includes a description of the relative activity and insertion context of the various TE superfamilies. Furthermore, the high transposition rates observed indicates that endogenous TEs can be used for saturation mutagenesis in grapevine. Such non-transgenic mutant vine populations would serve as valuable resources for both functional genomics studies and industry crop improvement.

8.2 Introduction

Since the work of Barbara McClintock, researchers have made use of both endogenous and transgenic TEs as tools for plant mutagenesis (see 2.2.8). Aimed at characterising gene function, these projects have generally made use of specific class II element families known to be highly active. The instability of new insertions arising from these elements frequently leads to genetic reversions, allowing confirmation of phenotypic associations. In the case of transgenic protocols, foreign TEs introduce a unique tag at each insertion site, which greatly aids the mapping of new mutations.

Although useful in functional genomics studies, transgenic and unstable TE insertions limit the economic value of any phenotypic improvements achieved. However, the transgenic introduction of well-characterised TE families may be avoidable. As our understanding of transposon biology has improved, an increasing number of active and recently active TE families have been discovered within plant genomes. In addition, radical improvements in high-throughput sequencing technologies and genome mapping have reduced the necessity for unique mutation tags.

8.2.1 Using next-generation sequencing technologies to detect TE mutants

Transposon-enriched sequencing involves the selective sequencing of genomic regions associated with known TE sequences. This has been achieved by first performing PCR amplification of TE-flanking regions (FUKAI *et al.* 2012; URBAŃSKI *et al.* 2012), or by tagging TE sequences with biotin and capturing these on streptavidin-coated matrices (GABRIEL *et al.* 2006; WILLIAMS-CARRIER *et al.* 2010). The reduction in overall complexity associated with sequence enrichment, together with multi-dimensional pooling strategies, increases the sample capacity for analysis, such that several hundred individual genomes can be analysed simultaneously (FUKAI *et al.* 2012; URBAŃSKI *et al.* 2012).

Although powerful, such an approach has same key drawback as S-SAP; that is, a bias is introduced by genotyping only pre-defined elements (see 7.5.4). Ideally, empirical data is needed to determine which elements are active within the genome before such enrichment strategies are used, in order to maximise the number of polymorphisms that are identified. Whole-genome shotgun sequencing is particularly suited to addressing this issue, as it offers the throughput to resequence entire genomes without enrichment.

A whole-mobilome analysis of real-time transposition in grapevine has not been undertaken before. Therefore, I chose to perform whole-genome resequencing on regenerated vines that were known to contain new TE insertions (see 7.5.3). Four pooled DNA samples were sequenced using Illumina paired-end sequencing. Each DNA pool was comprised of genomic DNA from five vines regenerated from EC exposed to the same conditions. To ensure that pre-existing transposon insertions were identified and eliminated from the data, two controls were used. Genomic DNA was extracted from a

sample of the EC cultures used to regenerate these vines, as well as from leaves of the parental vine from which the EC cultures were initiated. These two samples were also sequenced.

8.2.2 Identifying transposon polymorphisms in shotgun sequence data

In order to interpret, compare and search the wealth of sequence data generated by NGS technologies, a computational approach is necessary. As these data have become more common, a host of software tools that aim to identify TE insertions in genomic sequence data have been published. These have been fastidiously reviewed elsewhere (BERGMAN AND QUESNEVILLE 2007; LERAT 2009).

Among these, the ReAS software program (LI *et al.* 2005) takes advantage of the repetitive nature of mobile elements to define TE families. The software takes those sequences that are over-represented within whole-genome sequence data, and uses them to generate consensus sequences. Jaillon and colleagues (2007) used this tool to identify and define grapevine TE families during the *de novo* assembly of the grapevine genome. Another tool, the RetroSeq program (KEANE *et al.* 2013), searches alignment files generated from paired-end Illumina whole-genome sequence data mapped to a reference sequence to identify TE insertions not present in the reference.

To conclude this study on the activity of endogenous grapevine TEs, we used RetroSeq to identify all new TE insertion events in the vines regenerated from EC cultures. For the discovery phase of the analysis, the Repbase library of *V. vinifera* TE definitions, produced by ReAS as described above, was used to identify TE insertions. TE insertions in the genomes of regenerated vines were compared to those in the DNA of EC culture material and the parental vine so that pre-existing insertions could be removed from the analysis.

8.2.3 General calculations for saturation mutagenesis in grapevine

The term “saturation mutagenesis” describes the production of a population of individuals that contain representative mutations distributed across the genome (see 2.1.5). While such projects have proven useful to both science and industry, their value with regards to any particular trait depends whether or not an individual can be found that displays a corresponding genotypic and phenotypic change. In general, this depends on the likelihood of a mutation event at any given locus, known as the degree of saturation.

The chance of finding a specific mutant locus within a randomly mutagenised genome is dependent on the number of loci in the genome and the number of mutations. This relationship is expressed using the general probability equation below (1), where *p* is the probability of finding a given mutation, *l* is the number of genomic loci and *M* is the total number of mutations.

$$p = 1 - \left(\frac{l-1}{l}\right)^{M} \qquad (1)$$

The total number of new mutations (*M*) in a population can be expressed as a product of the number of individuals within the population (*n*) and the average number of mutations per individual genome (*m*):

$$p = 1 - \left(\frac{l-1}{l}\right)^{nm} \qquad (2)$$

From the equation it can be seen that the probability of detecting a mutant allele for any particular locus increases as both the size of the population and the rate of mutagenesis increase. However, complete saturation of a genome with random mutations is statistically impossible in a finite population. By solving for population size, the relationship can be expressed as:

$$n = \frac{log_{\left(\frac{l-1}{l}\right)}(1-p)}{m} \qquad (3)$$

Targeting every base-pair of a genome for mutation would require an impractically large population. Therefore, the haploid genome size (*G*) is divided at a defined resolution (*r*) into a number of loci (*l*):

$$l = \frac{G}{r} \qquad (4)$$

For example, the 487.1Mb genome of grapevine (JAILLON *et al.* 2007) can be divided at a resolution of 1kb into 487,100 loci. Equation (3) can then be used to determine the population size necessary in order to find at least one mutation in each 1kb locus at any confidence level (*p*), given a known mutation rate (*m*; Figure 8.1).

An ideal target for saturation mutagenesis projects is to achieve one mutant gene per individual. By defining target loci as individual genes, of which there are 29,927 in grapevine according to current estimates (http://plants.ensembl.org/Vitis_vinifera, retrieved 10th Oct 2013), the same equation (3) can be used to calculate the population sizes needed to achieve saturation mutagenesis of the genic space, based on a single mutant gene per individual. This special case is also plotted on Figure 8.1 (red line).

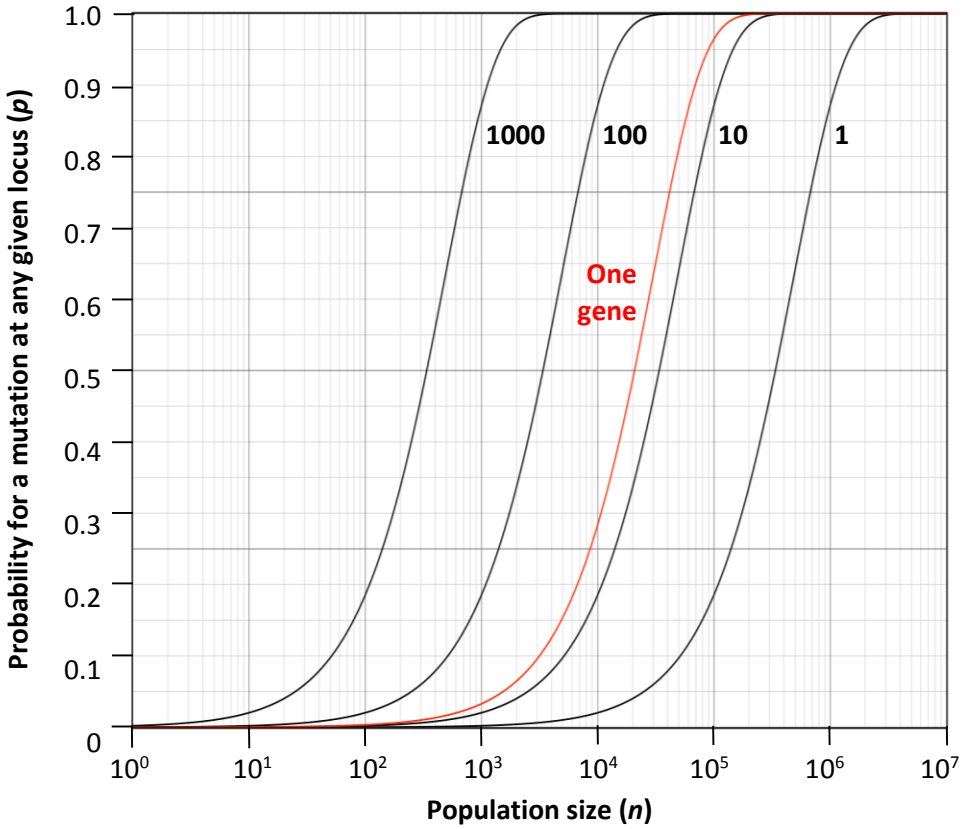


Figure 8.1 The degree of mutation saturation (*p*) within a population is controlled by both population size (*n*) and the rate of mutation (*m*). Calculations are based on a 1kb locus size. Mutation rates (*m*) are indicated next to the relevant curves. The red curve indicates mutation saturation given a mean mutation rate of one mutant gene per genome.

From this figure, it can be seen that a mutation rate of 100 new insertions per genome allows approximately 85% saturation of each 1kb locus with a population of 10,000 plants. However, the required population size increases to around 100,000 at a mutation rate of 10 per genome. A system producing new mutations at the ideal rate of one mutant gene per individual would include a mutant allele for 90% of all genes in a population of approximately 60,000.

8.2.4 Characterising the mobilome

Unique TE insertions in the regenerated vines were used to characterise the degree of transposition that has occurred during the process of embryogenic tissue culture. Since vines regenerated from EC exposed to biotic and abiotic stressors were included in the resequencing experiments, the relative activity of TEs in these vines could be compared to control vines regenerated from EC without stress exposure. These data are comparable with the transcript data collected earlier (Chapter 6), showing the positive effect of yeast cultures on TE activity in grapevine EC cells.

By overlaying the insertion loci data with publically available gene annotations for the grapevine genome, the context of TE insertions could be examined. This information was used to calculate location bias for various grapevine TE families, which is discussed in regards to TE insertion bias reported in other species.

The results presented here connect the various reports of epigenetic derepression in grapevine tissue following tissue culture and environmental stress events (SCHELLENBAUM *et al.* 2008; BARÁNEK *et al.* 2010) with the historical evidence of TE proliferation in the grapevine genome (Chapter 3), by showing that stress can raise the existing activity of multiple endogenous TE families in EC cultures. This has important implications for the concept of clonal identity in long-lived and vegetatively propagated species and also demonstrates that endogenous TEs can be used to produce new phenotypes from elite grapevine clones.

8.3 Methods

8.3.1 Sample preparation and sequencing

Of the 32 samples of pooled genomic DNA tested by S-SAP (Table 7.1), four were selected for whole-genome sequencing. Since the intention of this work was to capture and identify TE families which have become active, vines possessing either genotypic polymorphisms (identified by S-SAP) or phenotypic abnormalities were prioritised for whole genome sequence analysis. Three pooled DNA samples that had shown polymorphic S-SAP profiles, and included plants that displayed phenotypic abnormalities were selected. For simplicity, these were renamed as follows, Control 2: vines 16-20 regenerated from unstressed (control) EC cultures; Yeast 2: vines 6-10 regenerated from EC treated with *Pichia kluveri*; 50mM NaCl: vines 6-10 regenerated from EC treated with NaCl stress at 50mM (see 6.3.3 for stress treatment protocols). Neither genotypic nor phenotypic changes were previously detected in the fourth pooled DNA sample selected (Control 1: vines 6-10 regenerated from unstressed EC cultures).

Young leaf tissue from the five vines included in each sample pool was combined in equal amounts by weight and ground together in the presence of liquid nitrogen. Total genomic DNA was extracted from 500mg of the ground tissue, using the standard CTAB extraction protocol (DOYLE AND DOYLE 1990). DNA concentration was measured using a Qubit fluorometer (Life Technologies) and purity was assayed with a NanoDrop spectrometer (Thermo Fischer Scientific). The integrity of extracted DNA was assayed by running 100ng of each sample on a 1% (w/v) TBE agarose gel at 100V for 20 minutes.

For comparison, DNA was extracted from a 500mg sample of the embryogenic callus cultures that were used to regenerate the new vines and a 500mg sample of young leaf material from the vine used to initiate these cultures. Both of these samples were also processed as described above.

Genomic DNA shearing, library preparation and sequencing were performed by the Australian Genome Research Facility (http://www.agrf.org.au). Illumina paired-end library preparation was performed for each of the four samples using the TruSeq DNA Sample Preparation Kit (Illumina). The four libraries from the regenerated vine DNA were combined and sequenced in two lanes of a flow cell using an Illumina HiSeq2000 instrument. The parental leaf and embryogenic callus DNA samples were sequenced by NZ Genomics Ltd (http://www.nzgenomics.co.nz) using the same protocol. Each of these samples was sequenced using one full flow cell lane.

8.3.2 Transposable element identification

Raw sequence data processing, read mapping and TE identification were performed by Susan Thomson (Plant and Food Research NZ, Lincoln). Adapter sequences were removed from individual reads using fastq-mcf (https://code.google.com/p/ea-utils/wiki/FastqMcf) and the first 10 bases of each read were removed by hard-trimming in order to eliminate lower-quality regions. Read quality was checked using FASTQC before and after trimming (ANDREWS 2010). Reads were then mapped to the 12X PN40024 grapevine reference genome using Bowtie 2 (LANGMEAD AND SALZBERG 2012).

The RetroSeq software program (KEANE *et al.* 2013), was used to identify TE insertions in the mapped sequence data not present in the reference genome. The software identifies read pairs for which only one read can be mapped to the reference sequence and the other is maps to a database of repetitive elements (i.e. pairs which flank an insertion site junction). The collection of *V. vinifera* TE consensus sequences contained in Repbase Update (JURKA *et al.* 2005) was provided to as a reference database of TEs. Retroseq then clusters all of the read pairs associated with identified insertion sites based on genomic location and strand in order to identify the insertion site. Insertion site calls with a read depth greater than 5 and similarity greater than 90% to reference TEs were used for further analysis.

8.3.3 Comparing novel insertions

Sample identity tags were appended to the TE calls produced by RetroSeq. Insert location data for each of the samples (regenerated vines, EC culture and parental vine) were sorted according to chromosome location. RetroSeq TE calls that appeared within a 500bp window on multiple strands were excluded to remove pre-existing TEs.

Location data for the remaining insertions (i.e. insertions unique to one pool of regenerated vines) was combined with *V. vinifera* annotation data for predicted gene, mRNA and CDS sequences (available from www.genoscope.cns.fr) using Galaxy (GOECKS *et al.* 2010) to determine the context of each insertion. Unique TE loci were visualised as annotation tracks on the grapevine reference genome using Geneious (version 6; www.geneious.com). Counts of insert type and genomic context were produced for each RetroSeq dataset using Galaxy, and graphs were generated from these data using Excel (version 14; Microsoft Corporation).

8.4 Results

8.4.1 Illumina sequence data quality

Genomic DNA extractions produced around 20-30µg of high quality (260/280 = 1.8 - 1.9), intact DNA. Sequencing across two Illumina HiSeq2000 lanes produced approximately 83-95 million 100bp read pairs per sample. Read quality was high, with median Phred scores greater than 30 across all bases, except the first three bases in the case of the parental leaf tissue sample (Appendix B.7). These were removed during hard-trimming of the sequences.

8.4.2 Whole-genome sequencing (WGS) confirms transposition of four retrotransposon families

Novel polymorphisms were observed for each of the four retrotransposon families studied by S-SAP within the regenerated vines, as reported in Chapter 7. Between one (*Cremant*) and seven (*Gret1*) unique insertions per family were identified in the four sequence data sets (Figure 8.2). Although only five plants from each stress treatment were resequenced, it is notable that the fewest insertions (between one and three) were observed in vines regenerated from unstressed EC cultures (controls 1 & 2), while the most new insertions (eight) were observed in the vines regenerated from EC co-cultivated with yeast. These results reflect the stress (especially biotic) stimulation seen in the transcript data for these same retrotransposon families (see 6.4.3).

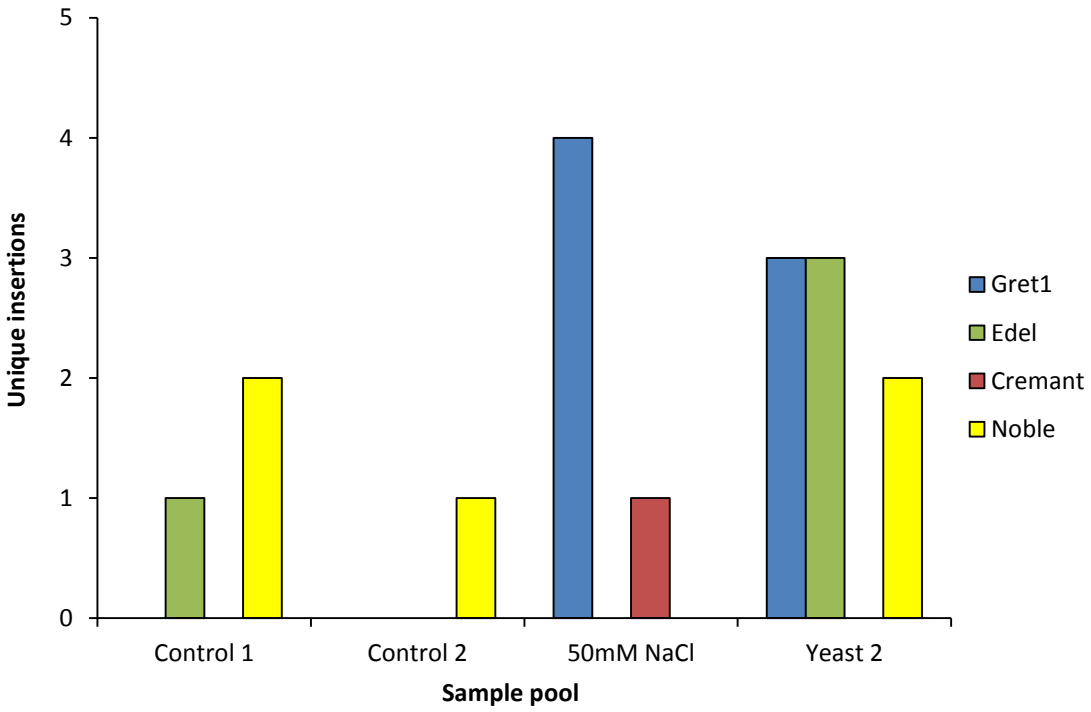


Figure 8.2 Unique insertions identified by WGS. Only the four retrotransposon classes previously analysed by S-SAP are shown.

The *Noble* elements, which were found to be expressed at a high rate in all samples in the EC cultures (see Figure 6.12) are represented by a few (one or two) new insertions in three out of four samples. In contrast, *Gret1* showed the highest relative transcription change in response to biotic stress exposure (excluding *Cremant*, which increased the least in absolute terms; Figure 6.7 and Figure 6.12) and accounts four three to four new insertions in each of the samples from stress-treatments but none in the control samples. The insertions of these four retrotransposon families were distributed across sixteen of the nineteen chromosomes (Figure 8.3).

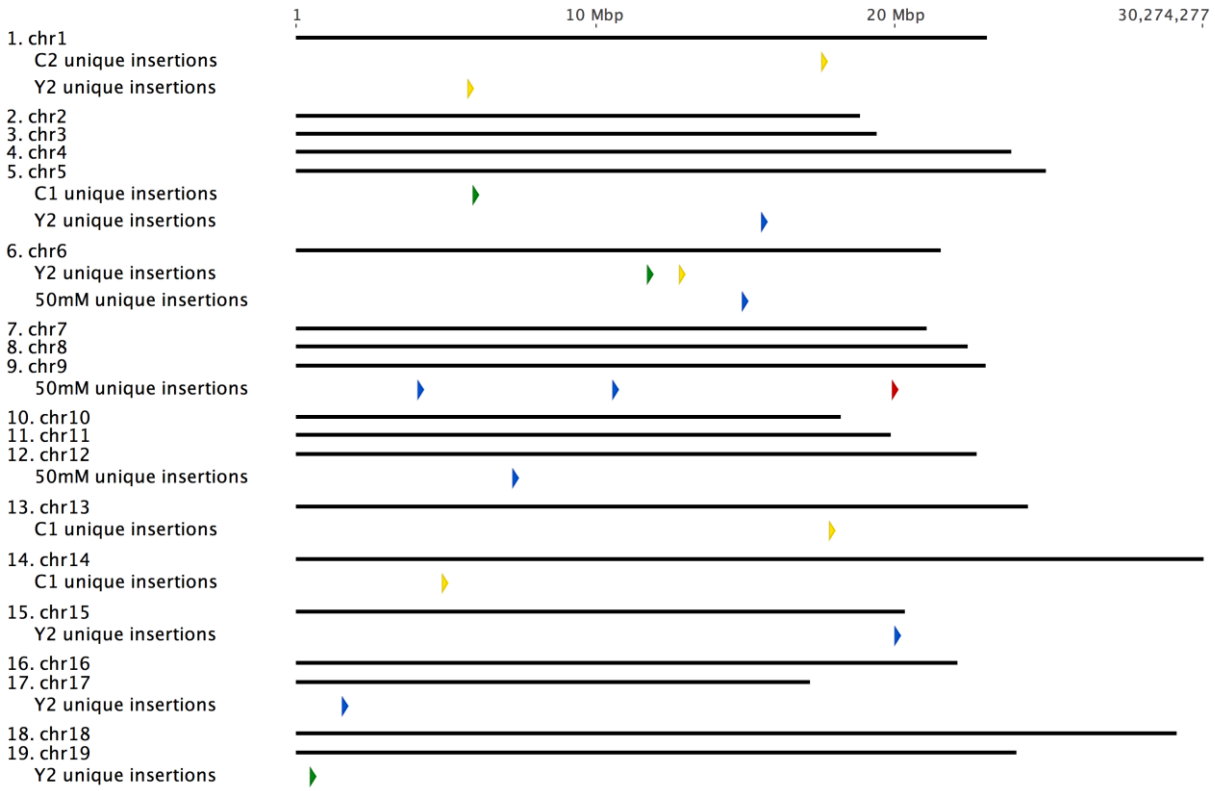


Figure 8.3 Unique insertions of four retrotransposon families identified by RetroSeq analysis of whole-genome sequence data. Mapped positions are shown relative to the pN40024 reference sequence. Black bars represent the 19 grapevine chromosomes. Retrotransposon families are coloured with the same pseudo-colours used for S-SAP analysis; blue: *Gret1*, green: *Edel*, red: *Cremant*, yellow: *Noble*. Unmapped chromosomal regions (including chr17_random, which contains a single Edel insertion) are omitted for clarity.

8.4.3 Numerous transposon polymorphisms are apparent within regenerated vines

RetroSeq comparison of alignment files of the resequenced vine genomes to the PN40024 *V. vinifera* reference genome identified a total of 19,305 TEs not present in the reference data. Visual inspection of the data revealed that insertions present in all samples were occasionally called as different TE families (Figure 8.4). Therefore, to eliminate false positive calls, any proximal (within 500bp) insertions within multiple samples (regenerated vines and parental material) were removed from the

data. As expected, the majority of the TE insertions (95.4%) were not unique, representing pre-existing insertions in the Pinot noir UCD5 genome not present in the PN40024 reference genome. Nevertheless, 1,282 unique insertions remained which were present only in a single sample (i.e. one pool of five regenerated vines; Table 8.1).

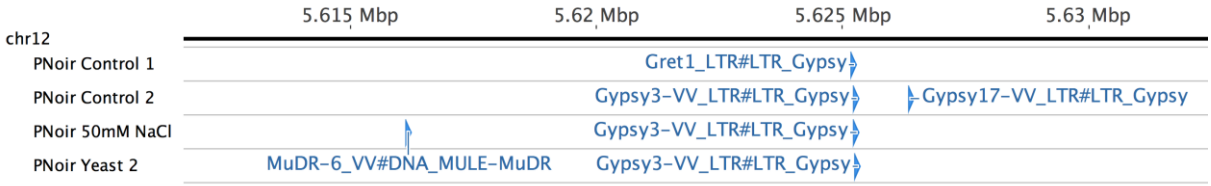


Figure 8.4 Identification of a conserved element. An insertion was identified by RetroSeq in each sequence data set within a 71bp region of chromosome 12. The element is called as *Gypsy3* in three tracks and as *Gret1* in the fourth. The canonical Repbase LTR sequences of *Gret1* and *Gypsy3* have a pairwise identity of 73%.

Table 8.1 Transposable elements not present in the 12X PN40024 reference genome. RetroSeq reports only TE insertions not present in reference sequence data. Those calls not present in the parental genotype or more than one regenerated samples were classified as unique.

Sample	Total RetroSeq calls	Unique inserts
Control1 (6-10)	4,356	244 (5.6%)
Control2 (16-20)	4,772	311 (6.5%)
50mM NaCl (1-5)	4,746	275 (5.8%)
Yeast 2 (6-10)	5,431	452 (8.3%)
TOTAL	*19,305*	*1,282 (6.6%)*

The S-SAP technique identified an average mutagenesis rate of 0.48 mutations per genome from the analysis of *Gret1*, *Edel*, *Cremant* and *Noble* retrotransposon families (see 7.5.3). In the resequenced data from twenty vines, a total of 17 new insertions were seen for these four families, equating to a combined mutagenic rate of 0.85 mutations per genome. The mutagenic rates of these element families, as measured by both techniques, are shown in Figure 8.5. The mutagenic rates of other TE families detected as unique mutations in the WGS resequenced samples are given for comparison.

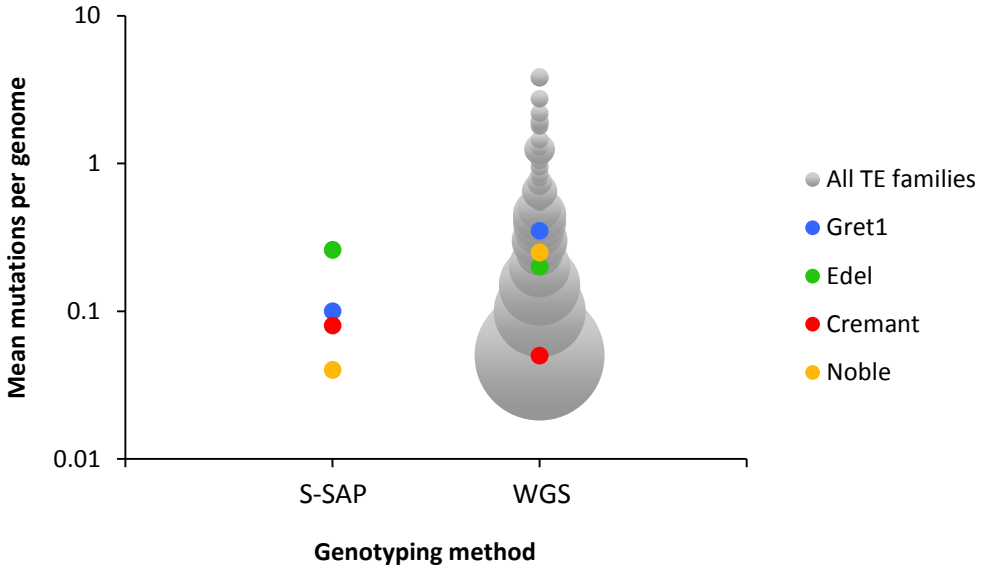


Figure 8.5 Mean number of new transposition events by TE family. Bubble size is indicative of the number of TE families active at an observed transposition rate. Transpositions rates of *Gret1*, *Edel* *Cremant* and *Nobel* retrotransposon families were measured using both S-SAP and WGS.

As Figure 8.5 shows, most of the TE families for which new insertions were identified produced fewer than 1 new insertion event per genome. In contrast, a number of families were much more active (those showing at least 1 new insertion per genome on average are listed in Appendix B.8, Table B.3). The highest rate of mutation was seen for *Gypsy-23* with an average of 3.9 transposition events per genome (79 unique insertions across all samples). The 1,282 total new insertion loci were dispersed across the nineteen grapevine chromosomes, with the majority of insertions falling within intergenic regions (Figure 8.6).

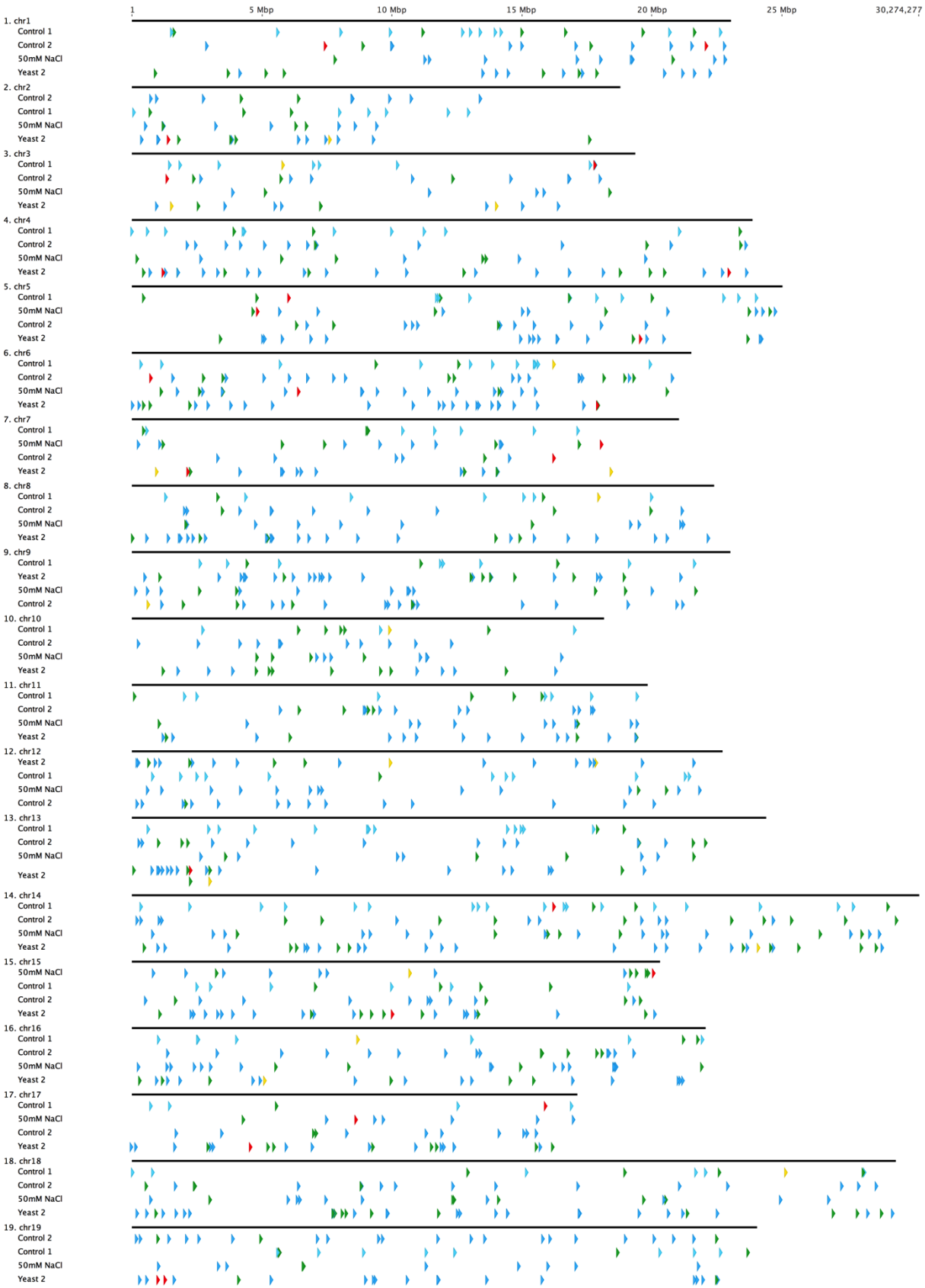


Figure 8.6 Novel transposon insertions in vines regenerated from embryogenic callus cultures. The mapped loci of all transposon insertions unique to a single pool of five regenerated plants are shown on the 19 chromosomes of the *V. vinifera* reference genome. The genomic context is indicated by colour as follows: red = within CDS, yellow = within UTR, green = within introns, blue = within intergenic regions. Unmapped chromosomal regions are omitted for clarity.

The number of new insertions was highest (452) in the pool of plants regenerated from embryogenic callus (EC) cultures that had been exposed to extracts from *Pichia kluveri* (yeast 2). However, approximately 250-300 unique insertions were also detected in each of the sequence data sets from plants regenerated from EC without stress exposure. Insertions were overwhelmingly, but not exclusively, found to be in intergenic regions, with only 5 to 11 insertions located within predicted coding sequences in each sample (Figure 8.7 and Appendix B, Figure B.12).

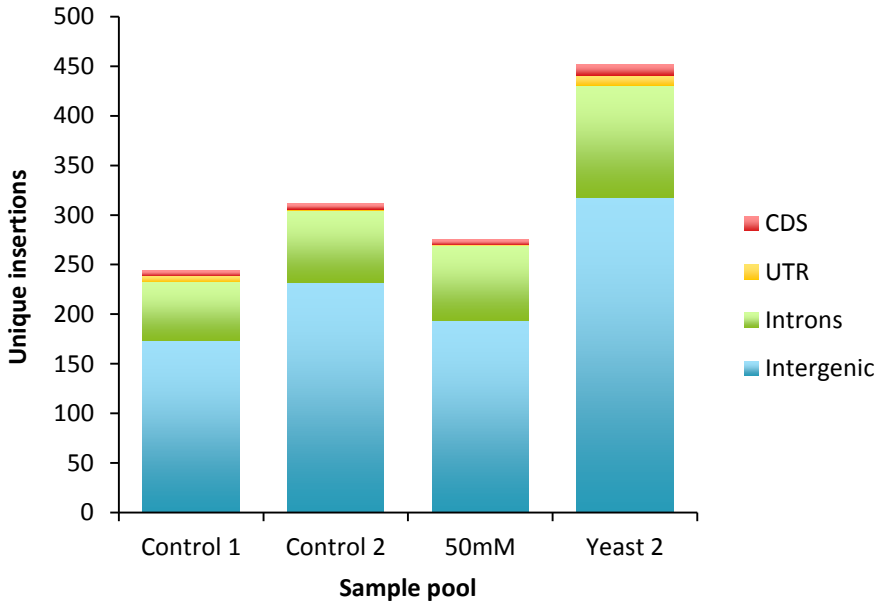


Figure 8.7 Numbers of new insertions by genomic context. Control 1 & 2: no stress exposure, 50mM NaCl: abiotic salt stress, Yeast 2: biotic *Pichia kluveri* extract stress

In each sample the majority of new insertions were associated with class I TEs, with Ty3-*Gypsy* elements accounting for the largest proportion of these (Figure 8.8). Class II elements accounted for between 37% and 45% of all new TE insertions, of which new *Mutator*-like elements (*MuDR*) were the most abundant.

The combined insertion data are summarised in Figure 8.9. This image shows the total activity of mobile elements in the 20 regenerated plants. New insertions are arranged by superfamily (right) and are linked to the genomic context (left) in which they were identified. The figure illustrates the insertion of most elements into the intergenic space, with a small number of insertions into the UTR and CDS regions.

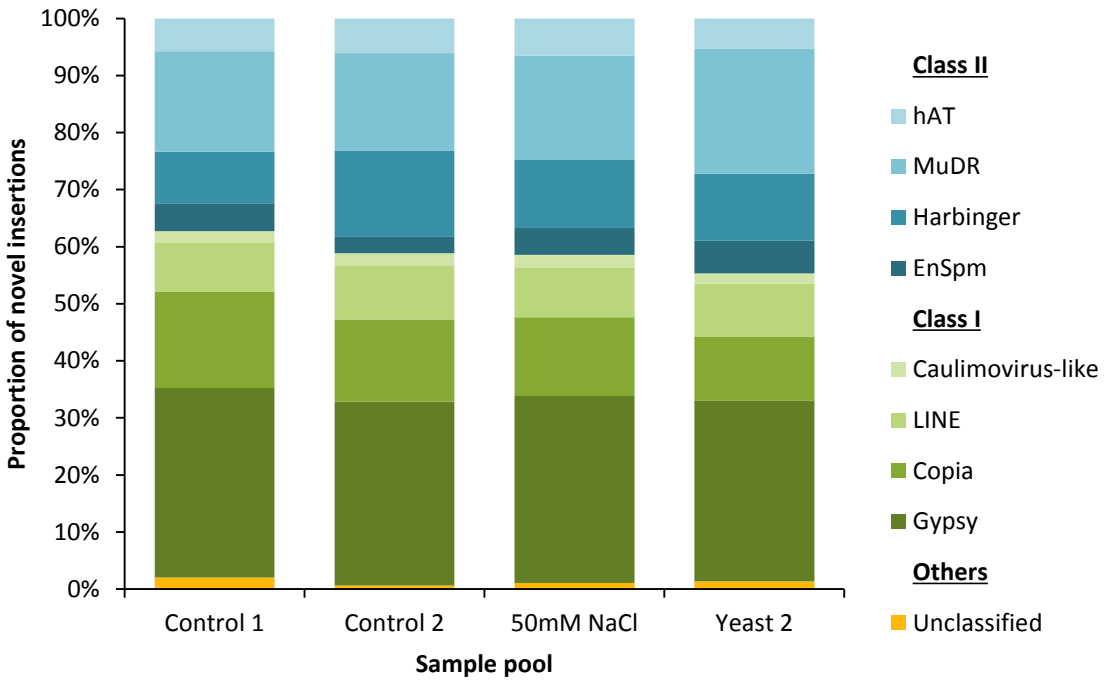


Figure 8.8 Relative contribution of TE types to new insertions. Control 1 & 2: no stress exposure, 50mM NaCl: abiotic salt stress, Yeast 2: biotic *Pichia kluveri* extract stress.

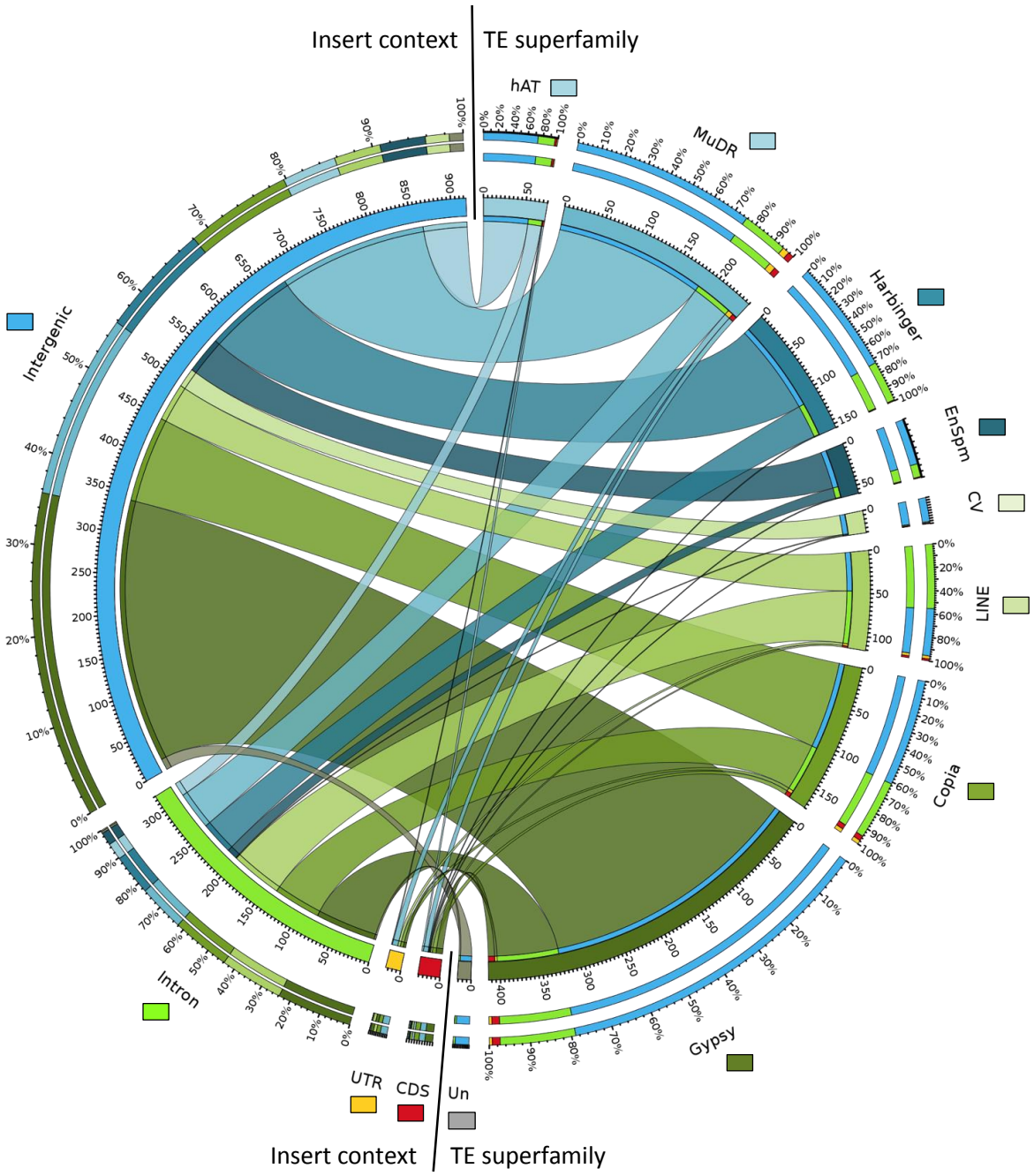


Figure 8.9 Total mobilome activity in the regenerated vines. Absolute numbers of new insertions in the 20 regenerated vines are displayed, with ribbons connecting the TE superfamily classes to which they belong (right) to the genomic contexts in which new the insertions were identified (left). Ribbons are coloured by TE superfamily, with class I TEs coloured green and class II TEs coloured blue. Outer rings indicate the proportion of each superfamily contributing to all new insertions (in the case of "Insert context") or the proportion of new elements ultimately inserting into each context (in the case of "TE superfamily". Un: unclassified TE families, CV: Caulimovirus-like, UTR: untranslated regions, CDS: coding DNA sequences. Graphic produced using Circos data visualisation software (KRZYWINSKI *et al.* 2009).

8.4.4 Location Bias

Although CDS-annotated regions comprise a minority of the grapevine genome (6.16%), fewer new TE insertions are present in these regions than would be expected by chance. The bias that is seen with respect to any genome region can be expressed as the ratio of the insertions detected to the number of insertions expected by chance. The number of insertions expected to fall in any genomic region by chance is the total number of insertions multiplied by the proportion of the genome included in the relevant region:

$$Bias = \frac{insertions_{observed}}{insertions_{total} \times genome\ proportion} \quad (5)$$

Using this approach, the overall location bias of the new TE insertions into each type of DNA (intergenic, intron, UTR and CDS) was calculated (Figure 8.10). An average score of 33% for TE insertions within coding sequences indicates a strong negative bias (i.e. only one third of the number of elements expected by chance are seen). The data also show a moderate negative bias towards UTR and introns (75% and 93%, respectively) and a slight positive bias (110% average) towards insertions within the intergenic space.

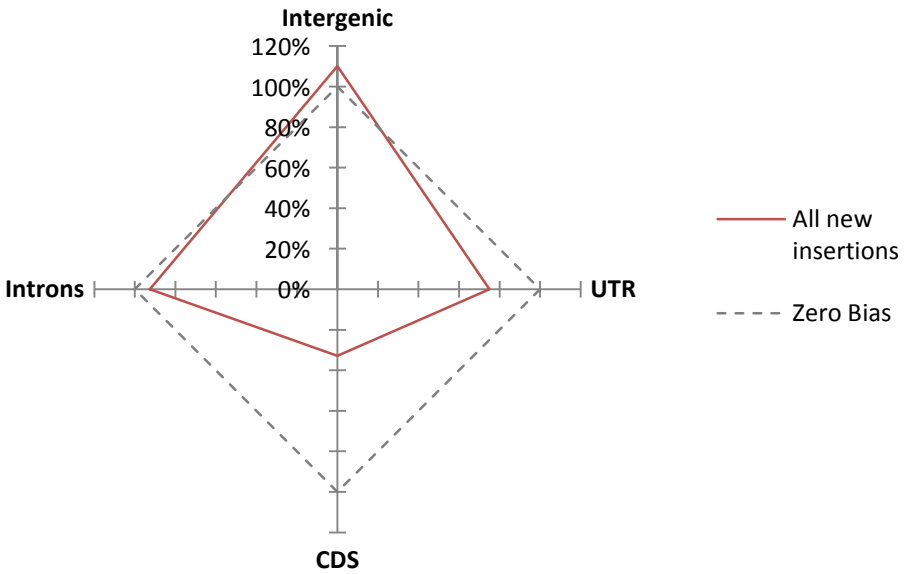


Figure 8.10 Location bias by sample. The total number of insertions in each genomic context is shown, normalised to the genomic coverage of the respective regions, so that a score of 100% represents zero bias with respect to the relevant genomic context (i.e. the number of insertions that would be expected by chance).

When location bias was analysed by superfamily (Figure 8.11), all showed a strong negative bias towards coding sequences. No *Harbinger*-like or *En/Spm*-like elements were identified in UTR regions, although more *Mutator*-like (*MuDR*), LINE and Ty1-*Copia* elements were found in this context than predicted by chance. Interestingly, more LINE transposons were found within introns than within the more abundant intergenic spaces (64 versus 47), resulting in a high location bias towards this context.

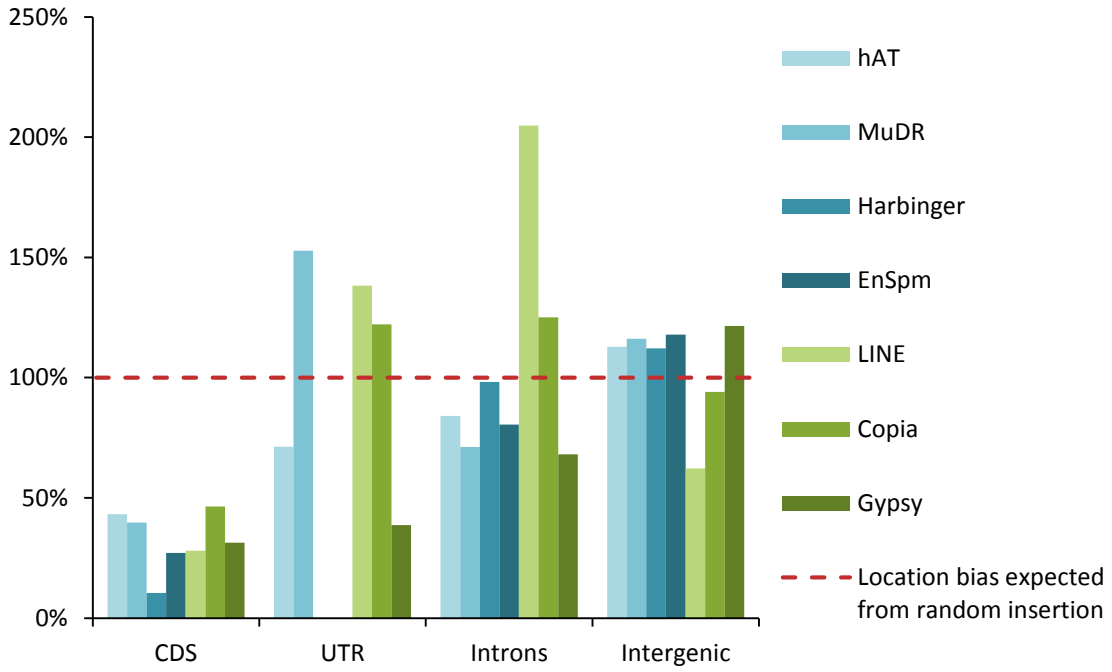


Figure 8.11 Location bias of new insertions by transposon superfamily. Insertion numbers are normalised to the relative genomic coverage of each feature type. A value of 100% (dashed red line) represents no location bias. TE superfamilies with fewer than 50 new insertions are excluded. Class I element superfamilies are coloured blue; class II element superfamilies are coloured green.

8.5 Discussion

8.5.1 Retrotransposon polymorphism among vines regenerated from stressed embryogenic callus

Whole genome sequence data showed that the novel transposition events detected by S-SAP in the regenerated vines (see 7.4.4) represent a very small proportion of the overall transposon activity. When the entire mobilome of 20 regenerated vines was analysed, 1,282 unique insertions were detected, which equates to a mean TE mutagenesis rate of 64 new insertions per vine.

Vines regenerated from EC stressed with salt (50mM NaCl) treatment showed an insertion number similar to those regenerated from control cultures, but vines regenerated from EC stressed by exposure to yeast extract showed 63% more insertions than controls. A comparable stimulation of retrotransposons was previously seen in the transcriptome of Pinot noir EC cultures following stress response (see 6.4.3). Cultures co-cultivated with six yeast strains for two days showed greatly increased levels of retrotransposon RNA, but those exposed to abiotic stresses showed no significant difference to the controls. From the data presented here, it can be concluded that the increased transcription of retrotransposons in yeast-stressed grapevine embryogenic culture cells ultimately leads to a higher incidence of transposition, which can be detected in adult vines regenerated from this tissue.

Two ways have been identified by which environmental stresses can increase the activity of plant TEs (discussed in 2.2.5). The first of these is the interaction of host transcription factors with TE regulatory motifs. The LTR regions of retrotransposons frequently harbour *cis*-regulatory element (CRE) sequences common to the promoter regions of stress response genes, including those known to increase transcription in response to mechanical wounding, light, pathogen attack and tissue culture (GRANDBASTIEN 1998; TAKEDA *et al.* 1999; VUKICH *et al.* 2009; WOODROW *et al.* 2010). Through their transposition, TEs rearrange these motifs within the genome. Their insertion can alter the expression patterns of target sites, occasionally to provide an adaptive advantage (TAKEDA *et al.* 1999; NAITO *et al.* 2009). In this way, thousands of TEs have become domesticated, fulfilling regulatory roles in their host genomes. Within five grapevine retrotransposon families analysed, between 5 and 68 stress-related CREs were identified per LTR sequence (see 3.4.3) indicating that these elements are likely to be transcribed under stress conditions.

However, transposons are generally highly methylated in plant genomes, which prevents access to their LTR regions by proteins involved in transcription. The second route by which endogenous TE activity is increased is the release of epigenetic silencing. Recent work has shown that epigenetic modifications that prevent the transcription of transposable elements can be altered by abiotic stress events and biotic stressors in *Arabidopsis*, resulting in a derepression of these elements (PECINKA *et*

al. 2010; DOWEN *et al.* 2012). After the stress period has passed, a siRNA biogenesis-dependant silencing pathway restores the epigenetic repression (PECINKA *et al.* 2010; ITO *et al.* 2011). Epigenetic derepression of the mobilome therefore allows a more general mobilisation of multiple TE families.

Although research into the epigenetic modification of grapevine tissue by environmental changes is sparse, evidence of methylation changes following temperature stress have been detected in Müller Thurgau and Riesling (BARÁNEK *et al.* 2010). Furthermore, Schellenbaum and colleagues found that vines of the Syrah and Chardonnay varieties regenerated by somatic embryogenesis showed epigenetic changes (2008). There is currently no direct evidence for epigenetic change in the vines produced during the course of this project. However, taking into account these previous reports, the elevated TE transcript profiles (see 6.4.3) and the diversity of new insertion types observed in regenerated vines (Figure 8.9), it is probable that the process of somatic embryogenic tissue culture results in broad epigenetic modifications to the grapevine genome, which are accompanied by a general increase in mobilome activity. Our research group is currently beginning to investigate this hypothesis using whole-genome methylation data from bisulphite sequencing of grapevine EC cultures and control samples.

Other environmentally responsive mechanisms of TE mobilisation may also contribute towards mutagenesis in grapevine EC cultures. An example is the temperature-dependant nuclear import of the transposase of the *Tam3* transposon in Antirrhinum (FUJINO *et al.* 2011).

8.5.2 Active element types

Despite representing less than 2% of all ancient insertions in the grapevine reference genome (JAILLON *et al.* 2007; VELASCO *et al.* 2007; BENJAK *et al.* 2008), class II elements proved to have a disproportionately high mutagenic capacity, accounting for almost half (42%) of the new insertions detected in the regenerated vines. Among this class, members of the *Mutator*-like (*MuDR*) superfamily of elements showed the most mutagenic potential, accounting for 19% of all new insertions (12 new insertions per plant on average). Elements of this type are transcriptionally active in a wide variety of plant species and are known to cause novel mutations in maize, rice, *Arabidopsis* and the fungus *Fusarium oxysporum* (LISCH *et al.* 1995; SINGER *et al.* 2001; CHALVET *et al.* 2003; GAO 2012). Their activity has previously been used to create mutagenised populations in maize (WALBOT 1992b).

Both the Ty1-*Copia* and Ty3-*Gypsy* superfamilies of class I elements were highly represented among new insertions (14% and 32% respectively). Carrier and co-workers (2012) have also reported high polymorphism of Ty3-*Gypsy* elements in genomic sequence data from Pinot noir clones, using alternative software tools. Contrary to expectations, the four retrotransposon families selected for

study based on genomic evidence of their recent mobility (*Gret1*, *Edel*, *Cremant* and *Noble*; see Chapter 3) were not found to be among the most mutagenic TE families. The *in silico* analysis of the integrity and past transposition patterns of individual TE families (Chapter 3) was therefore not predictive of the elements that are currently most active under the stress conditions tested. This may be due to the pattern of proliferation that TEs follow within a genome, which involves bursts of transposition followed by long periods of inactivity (KIDWELL 2002). Transposons that were recently active, and of which many intact genomic insertions can therefore be found, may in fact be those against which the host silencing systems are currently most attuned and which are at the beginning of the long interval of dormancy that separates such bursts of activity.

In addition to the Ty1-*Copia* and Ty3-*Gypsy* superfamilies, LINE retrotransposons were well represented among the new mutations (9% of all new insertions), including the family for which the second most new insertions were identified (*VLINE3_VV*). This family is one of eleven non-LTR retrotransposon families that have been identified in the grapevine genome (www.girinst.org/repbase/). The LINEs are abundant in plants (SCHMIDT 1999) and individual elements have, as for many other TE types, been shown to be activated in response to environmental stresses (VISIOLI *et al.* 2013). Two active plant LINE elements (*Karma* in rice in and *Lib* in sweet potato) have previously been observed to cause new mutations (KOMATSU *et al.* 2003; YAMASHITA AND TAHARA 2006).

Because the Repbase Update collection of grapevine TEs was used to annotate new TE insertions in this study, it is possible that the high number of *VLINE3_VV* mutations recorded is related to the low number of defined grapevine LINE elements; that is, that the LINE retrotransposons of grapevine are less well defined as individual families. Similarly, the multiple retrotransposon families for which very few new insertions were observed (see Figure 8.5) might in fact be miss-identified insertions belonging to a family of high sequence similarity. The RetroSeq software was occasionally inconsistent in the identification of ancestral TE insertions where multiple families have similar LTR sequences (Figure 8.4). Sequence comparison of the new insertions, while beyond scope of this project, should be able to clarify these issues.

8.5.3 Insertion site bias

A prominent bias was observed against the identification of new elements within gene coding sequences (Figure 8.10). Insertions into CDS regions are most likely to disrupt the synthesis of proteins, and therefore have the greatest chance of effecting phenotypic change. As with all mutations, the great majority of those that alter phenotype are expected to be deleterious. Lethal mutations could arise from the insertion of TEs into genes that produce proteins responsible for any one of a myriad of critical biochemical pathways, or into the genes of transcription factors that regulate their expression. Disruption of any genes necessary for the cell division, growth or

differentiation of EC cells would prevent the regeneration of vines carrying such alleles. Likewise, mutations that are not lethal, but are still detrimental to the fitness of an individual, may prevent the ultimate survival of the young vine from tissue culture through hardening. The lack of insertions within coding sequences may therefore not be mainly due to an insertion bias of the transposon families, but rather to the negative selection that limits the survival of EC cells with CDS insertions during germination and hardening.

A less pronounced overall bias was also visible with respect to UTR and intron sequences. These genic regions house motifs responsible for regulating gene expression and are therefore also more likely than intergenic space to be associated with phenotypic change. The large size of many TEs means that insertions within introns may disrupt gene transcription (LUEHRSEN AND WALBOT 1992). In contrast to this general trend, the Ty1-*Copia*, LINE, and *MuDR* (*Mutator*-like) superfamilies showed insertion biases of 122%, 138% and 153% within UTR sequences respectively, which are higher than would be expected by chance. This increased bias in spite of negative selection pressure is a possible indication of insertion bias towards this context for certain elements of these superfamilies. It is important to note, however, that since UTR regions comprise the smallest context in the genome (approximately 2%), the sampling error is expected to be highest with respect to these regions. Although no studies have yet analysed insertion site biases of these TE families in grapevine, a bias towards 5' UTR sequences has been previously reported for *Mu* elements in maize (DIETRICH *et al.* 2002).

Insertion bias has also been previously reported for the LINE retrotransposons. In Drosophila, these elements are known to target telomeric regions, fulfilling the role of telomerase absent in this species (PARDUE *et al.* 2005). There is also some disagreement about whether the mammalian L1 nuclease has an insertion bias towards GC-poor regions, as a result of the enzyme's TT/AAA recognition site (OVCHINNIKOV *et al.* 2001; GRAHAM AND BOISSINOT 2006). Insertion bias of LINE elements has not yet been reported in plants, however in this study more than twice as many LINE transposons (205%) were found within introns as would be predicted under conditions of random insertion. This constituted the strongest evidence of insertion bias within the resequenced data. No clustering of LINE insertions at telomeres was observed.

It may be that those elements that show a positive location bias towards introns and UTR sequences, but a negative bias towards coding sequences (specifically *MuDR*, *Ty3-Copia* and LINE families) insert at a higher frequency into uncondensed euchromatic DNA, but experience negative selective pressure if the insertions disrupt protein synthesis.

8.5.4 The genomic impact of mobilome activation

Somaclonal variation describes the genotypic and phenotypic variation observed among plants regenerated from tissue cultures. Although this genomic instability is frequently reported as a negative by-product of *in vitro* propagation, it has been recognised as a source of new genetic material for crop improvement for over thirty years (LARKIN AND SCOWCROFT 1981). Known causes include pre-existing chimerism in explant tissue, chromosome number and structure variations, activation of transposable elements, single-nucleotide changes and microsatellite repeat number polymorphisms (BAIRU *et al.* 2011).

In this study, a high degree of genetic variation due to transposon activity was seen in vines regenerated from EC cultures. Active endogenous transposons have been linked to somaclonal variation in several other species including banana (MUHAMMAD AND OTHMAN 2005), barley (CAMPBELL *et al.* 2011), lotus (FUKAI *et al.* 2010), maize (PESCHKE *et al.* 1987; PESCHKE AND PHILLIPS 1991; BARRET *et al.* 2006), petunia (VAN HOUWELINGEN *et al.* 1998), rice (KIKUCHI *et al.* 2003), sugar cane (SUPRASANNA *et al.* 2010) and sweet potato (YAMASHITA AND TAHARA 2006). Unlike the data presented here, these previous studies used PCR-based techniques to identify polymorphic markers associated with specific TE families, so it is not possible to know whether the mobile elements identified are atypically mutagenic within these genomes, or only represent a fraction of the total mobilome activity.

Only two studies have so far used WGS to characterise genetic variation in plants regenerated from tissue culture. These studies found that somaclonal variation in *Arabidopsis* (JIANG *et al.* 2011) and rice (MIYAO *et al.* 2012) was largely due to SNPs and not TEs, although the *Tos17* retrotransposon of rice was seen to be active. In contrast, analysis of different clones of the Pinot noir variety has shown that the majority of interclonal genotypic polymorphism is associated with TEs (CARRIER *et al.* 2012). Although rates of SNP and indel polymorphism were not measured in the work presented here, the high polymorphic rates observed for multiple mobile element families demonstrate the mutagenic potential of the grapevine mobilome, which may explain the high rates of TE polymorphism between clones.

When comparing these species, however, it is important to consider the differences in the life cycles of these organisms. Unlike *Arabidopsis* and rice, grapevine has an agricultural history that involves millennia of vegetative propagation. The activity of mobile DNA elements is most actively repressed in gametic tissues (see 2.2.6). Therefore, it could be that by preventing the plant from passaging through this developmental stage, the periodic reinforcement of TE silencing has been lost, allowing certain mobile elements to become more active than they otherwise may have been. Transposition is also known to have been the cause of certain grapevine bud sports (KOBAYASHI *et al.* 2004; FERNANDEZ

et al. 2010). Consequently, centuries of bud sport selection by viticulturists may also have inadvertently contributed towards the propagation of clones that harbour active TEs.

These results have serious implications for the concept of clonal uniformity among populations of crops propagated or by tissue culture. Although somaclonal variation and somatic mutations have long been recognised as both a problem and an opportunity, these data show that the grapevine mobilome may posses a particularly high capacity for the generation of new genetic variation. Transposon-driven variance may, to a limited extent, alleviate the lack of genetic diversity that is caused by continuous vegetative propagation.

Of further interest is the degree to which TE mobility causes genetic change in transformation experiments. Somaclonal variation has been observed at the epigenetic, genetic and phenotypic levels in grapevines regenerated from cell cultures (SCHNEIDER *et al.* 1996; KUKSOVA *et al.* 1997; SCHELLENBAUM *et al.* 2008), although this has not yet been studied with respect to transgenic vines. Transposition events have the potential to disrupt or alter the expression of nearby genes, as discussed earlier (see 2.2.6). The significant mutagenesis caused by TE mobilisation in the 20 vines sequenced in this study indicates that any vines regenerated from embryogenic cultures are likely to possess a large number of novel genetic mutations, which may confound the phenotypic analysis.

8.5.5 Saturation mutagenesis using TE activity in embryogenic cultures

On average, 90 new TE insertions were observed per individual regenerated from yeast-stressed somatic embryos, 2 of which were within coding sequences. These values can be used to calculate the population sizes required for saturation mutagenesis by stimulating TE activity in EC cultures.

By substituting for genomic loci according to equation (4), equation (3) can be expressed as:

$$\boldsymbol{r = \mathrm{G}\left(1 - \sqrt[nm]{1 - p}\right)} \qquad \textbf{(6)}$$

Using this arrangement of the equation, Figure 8.12 illustrates the relationship between the resolution of mutant loci in a saturation mutagenesis experiment and the size of the population, at a probability of 95%. In general, as the population size (*n*) increases, the average number of base pairs between any two mutations (*r*) decreases in inverse relationship. The graph shows, for example, that at the observed rate of 90 mutations per genome there is a 95% chance that a population of ten thousand vines will have a mutation in any region of the genome approximately 1,500bp in length.

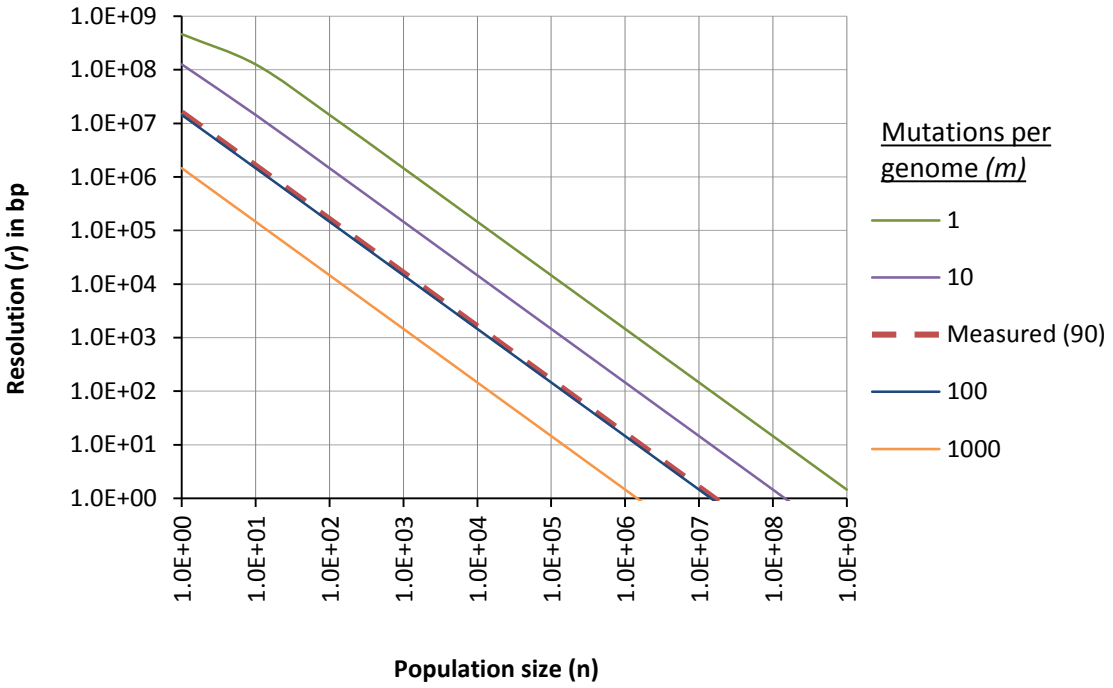


Figure 8.12 Mutation resolution decreases as population size increases (p = 0.95). The mean genomic distance between mutations within a population has an inverse relationship with the size of the population. Curves are plotted according to equation (6), using the genome size estimate of Jaillon *et al.* (2007). The observed average value of TE 90 mutations per individual genome (dashed red line) is shown against mutation rates of 1 to 1000 mutations per genome for comparison.

Since mutations in the intergenic spaces are less likely to have phenotypic consequences, a more appropriate estimation of mutagenesis can be achieved by only considering rates of mutation within genic regions. In the case of plants regenerated from yeast-stressed EC cultures, an average of 27 mutations within genes were observed, 2 of which were within coding sequences. Equation (3) can be used to calculate the population sizes needed to obtain mutation saturation within genes or specifically within the coding sequence of each gene (Figure 8.13). This graph shows that at average the rate of mutagenesis observed among the sequenced vines, a population of approximately 63,000 vines would represent a mutant saturation rate of 99% within coding sequences. At a planting density of 6,000 vines per hectare, such a population could be maintained in little more than 10 hectares of vineyard space.

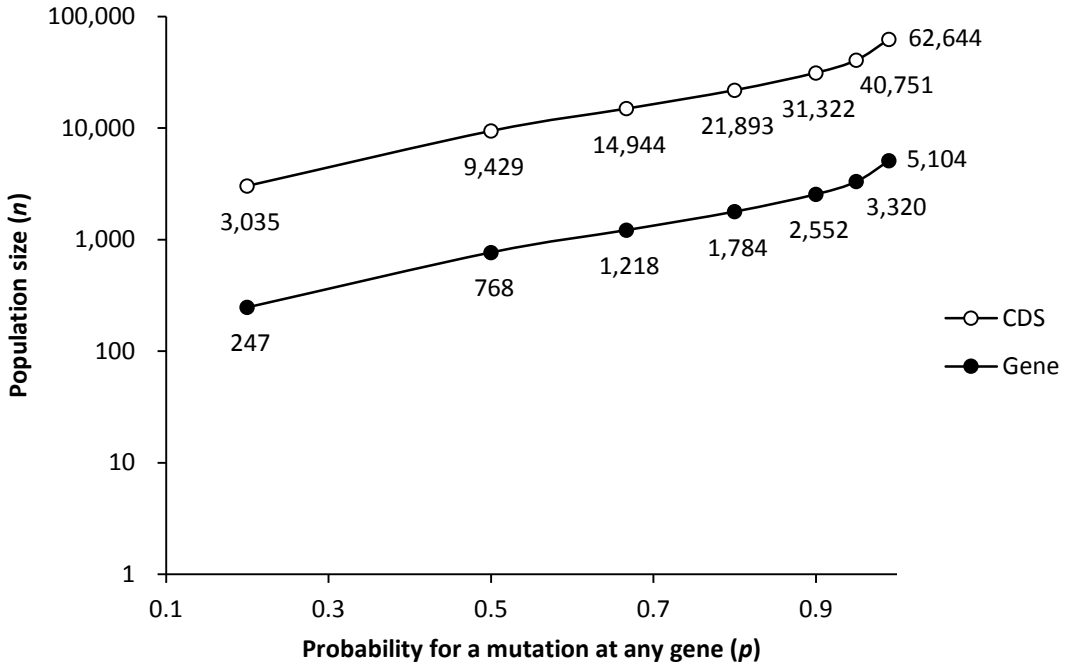


Figure 8.13 Population sizes required for saturation of the grapevine genic and CDS regions with transposable elements. Calculations are based on the average transposon mutation rate observed in the vines regenerated after biotic (Yeast 2) stress.

Although genes are commonly used as defined loci for saturation mutagenesis experiments (WALBOT 2000), it is important to note that these calculations do not account for variations in gene size. The probability (*p*) therefore relates to genes of average size. For genes smaller in size than the genomic average, the chance of the population containing a mutant allele is less than *p*. For genes larger than the average size, the chance is greater than *p*.

8.5.6 The epigenetic impact of transposition

In addition to the direct genetic changes caused the transposition of mobile elements, the epigenetic modifications that regulate gene expression are also influenced by the presence of mobile elements. Without directly measuring changes to the epigenome, it is not possible to comment on the epigenetic effect that extensive TE mobilisation has had on the regenerated vines produced. However, considering the diversity of elements that have shown transposition, it is likely that a general hypomethylation of transposon sequences has occurred in the EC cultures (see 6.5.4).

In addition to the need to regain control of the mobilome by re-establishing transposon silencing, regenerated vines are also faced with the challenge of aberrant transcripts produced from new TE

insertions. The likely effect of this can be speculated based on previous published work. Dowen and colleagues (2012) recently showed that a general activation of transposon-associated sequences in response to salicylic acid stress was linked to methylation changes in these regions in *Arabidopsis*. Furthermore, they noted that genes adjacent to these TEs were often also upregulated. The general activation of TE sequences in grapevine is therefore likely result in similar expression changes in nearby genes.

Transposon insertions have been linked to the maintenance of heterochromatin (see 2.2.6), and methylation targeted to TE insertions can spread to affect the expression of nearby genes (IIDA *et al.* 2004). Expression of aberrant transcripts from TE insertions sites also triggers host silencing systems, resulting in silencing of not only the site from which such transcripts originate, but also homologous sequences in *trans* (GIRARD AND HANNON 2008). Transposition therefore creates regulatory networks, affecting multiple sites across the genome and altering the epigenetic state of DNA not otherwise impacted by new insertions. The overall phenotypic impact of the numerous transposition events observed in the vines regenerated from stressed EC cultures during the course of this project probably results from extensive epigenetic as well as genetic changes.

8.6 Conclusions

The stimulatory effect of stress on the transcription of mobile elements (Chapter 6) ultimately results in an increased number of new transposon insertions. The polymorphism detected among four retrotransposon families by PCR-based assays (Chapter 7) represents only a small fraction of the total variation of the mobilome seen in vines regenerated from embryogenic callus cultures. The new polymorphisms were not the result of a small number of highly active TE families, but rather appear to be due to a broad activation of mobile elements, likely caused by a general release of transposon silencing.

This extensive mobilome activity shows that multiple element families retain the potential for transposition, which is in agreement with the high number of TE families for which complete copies found in the reference genome sequence and which are represented in grapevine transcript databases (Chapter 3). These results also demonstrate that an *in silico* analysis of the number and age of past insertions by family is not predictive of the most mutagenic elements that exist within the contemporary genome.

Despite a strong negative bias towards coding region insertions among the regenerated vines, the rate of endogenous transposon mutagenesis was such that high levels of genomic coverage (95-99% of coding sequences) can be achieved within feasible population numbers (40,000 – 65,000 vines).

Chapter 9

Overall Conclusions

9.1 Summary of findings

9.1.1 Recently active LTR-retrotransposons in grapevine (Chapter 3)

A broad overview of class I transposable elements was performed using the *V. vinifera* reference genome data. By comparing the multiple sequences corresponding to individual retrotransposon families it was shown that most families exist almost entirely as degraded elements. Others appear to have undergone recent bursts of transposition, and endure as multiple complete functional units, including intact open reading frames and identical long terminal repeats. The data presented disprove the stated hypothesis:

$H_{0\text{-}1}$: No evidence can be found for differences in the historical activity of transposon families within the grapevine genome sequence data (see 1.6.1).

Internal regions proved most useful for comparing retrotransposon families, while LTR sequences, which displayed greater variation, were preferable for comparing multiple insertions of the same family. A phylogenetic comparison of retrotransposon families between angiosperm genomes adds to the growing body of evidence for horizontal transfer of retrotransposons between species.

Within a selection of five families, stress-related response motifs were found to be common, indicating potential expression of these elements under extreme conditions. Evidence of TE expression was found to be abundant within EST databases, particularly from stressed tissues.

9.1.2 S-SAP is a rapid and reliable tool for genotyping TE polymorphisms (Chapter 4)

A variety of PCR-based techniques have previously been used to genotype grapevine clones and varieties, with those anchored to retrotransposon sequences showing the highest degree of polymorphism. In order to obtain a tool capable of identifying transposition with high sensitivity, the S-SAP technique was optimised for use in grapevine and tested on four specific grapevine retrotransposon families. By tagging amplification products with fluorescent chemical dyes and separating these using capillary electrophoresis, a resolution and sensitivity level was achieved that far supersedes gel-based analysis, allowing polymorphisms to be detected among the high numbers of monomorphic background elements. This approach allowed transposon profiles to be rapidly generated in a highly automatable manner, with a size-calibrated digital output that allows comparison of samples between experiments.

The analysis of 32 genotypes from 12 different varieties using this method revealed polymorphisms that distinguished species, varieties and clones, disproving the following hypothesis:

$H_{0\text{-}2}$: Phenotypically different vines grown from vegetative material have identical TE insertion patterns (see 1.6.2).

Once bands of interest were identified, these were isolated by repeating the specific amplification using a biotin tag and gel-based separation. The combination of these two systems means that most genotyping can be performed in a high-throughput manner, without the need for gel analysis. Only samples containing polymorphisms for which locus information is required need be visualised on a gel.

Approximately two thirds of the RBIP markers generated from flanking sequence data proved useful for rapid genotyping of genomic DNA, which could be performed directly on leaf tissue preserved on FTA cards. The generation of genotypic tests associated with specific TE polymorphisms disproves the null hypothesis:

$H_{0\text{-}3}$: Differences between TE insertion profiles cannot be associated with the presence or absence of a single TE insertion at a specific locus in the genome (see 1.6.3).

9.1.3 Improved leaf tissue penetration by aqueous suspensions for enhanced *Agrobacterium*-mediated transformation (Chapter 5)

To improve penetration of microorganism suspensions into grapevine vegetative tissue, an organosilicone surfactant was used to reduce the surface tension of aqueous media. To quantify the effect that this had on the interaction between microorganisms and plant tissue, this technique was applied to transient *Agrobacterium*-mediated transformation of grapevine leaf tissue. Transient tissue transformation is a fundamental tool of *in vivo* gene function assays, but generally has low efficiency in grapevine and requires the isolation of plant tissue in a vacuum. The addition of the surfactant greatly improved transformation efficiencies at ambient pressure, significantly improving the utility and simplicity of this important technique.

To quantify transformation efficiencies, a construct comprising the grapevine VvMYBA1 gene under control of a CMV 35S promoter was used. The expression of VvMYBA1 was found to be sufficient to induce anthocyanin pigmentation in leaf tissue of all varieties tested. The accumulation of anthocyanin was confined to individual cells and was visible under normal lighting. Consequently, this construct proved more precise than GUS as a reporter and allowed more accurate transformation scoring in a non-destructive assay.

9.1.4 TE expression in response to stress treatments (Chapter 6)

The expression of four grapevine retrotransposon families was shown to increase following exposure of cell cultures to environmental stress, with a particularly high response seen following co-cultivation with live yeasts. This response was specific to embryogenic tissue cultures grown on solid medium. The high variation in TE transcript levels in these stressed cultures disproves the null hypothesis:

$H_{0\text{-}4a}$: Transcript levels of grapevine LTR-retrotransposons are not altered by environmental stress events (see 1.6.4).

In plants, methylation is primarily associated with repeat elements. Work done by other researchers indicates that TE transcription in response to stress, as seen here, is a result of the demethylation of these elements (BARÁNEK *et al.* 2010; TITTEL-ELMER *et al.* 2010). The increased transcription of all four TE families tested in these experiments is consistent with that explanation.

9.1.5 A tagged mutant population of Pinot noir vines (Chapter 7)

A procedure was developed for the stress exposure and germination of embryogenic callus cultures. Vines regenerated from cells stressed with ten different treatments were germinated, hardened off and transferred to large pots, where they continued to grow until entering winter dormancy. Genotypic screening of the regenerated vines using the optimised S-SAP procedure revealed retrotransposon-associated polymorphisms, null hypothesis $H_{0\text{-}4b}$ is therefore rejected.

$H_{0\text{-}4b}$: Multiple grapevine plants regenerated from the same stressed tissue have identical TE insertion patterns (see 1.6.4).

9.1.6 A comprehensive assay of grapevine mobilome activity (Chapter 8)

Grapevine transposable elements have been previously identified and defined by amplification of the genome with degenerate primers (MOISY *et al.* 2008), by aligning highly repetitive whole-genome sequence data (JAILLON *et al.* 2007) and by performing searches on assembled reference genome data (BENJAK *et al.* 2008; BENJAK *et al.* 2009). These studies have made great strides in defining and characterising the elements that comprise the grapevine mobilome. Using a variety of molecular marker systems, specific transposons have been shown to be polymorphic among varieties, which has led to development of genotyping tools that can be used to discriminate varieties, and in some cases, clones (see 4.2.3). While these studies have helped to establish an understanding of the contribution that transposons have contributed to the historical genetic variation within *Vitis*, they are unable to comment on contemporary transposon activity.

This project has shown for the first time that multiple transposon families, from most TE superfamilies, retain the potential to become mobile and cause new mutations within the grapevine genome. New TE insertions were seen to increase following stress treatments, in accordance with the transcription data measured earlier (Chapter 6). The insertions showed a location bias against genic regions, particularly with respect to protein-coding sequences, which is likely to be due to negative selective pressure. Certain superfamilies did show evidence of insertion bias however, most notable of which was a prevalence of LINE insertions towards introns.

Interestingly, retrotransposon families that showed the most evidence of recent mobility (Chapter 3) were not responsible for the greatest numbers of retrotransposon insertions in the regenerated vines. Consequently, this study fails to reject the following null hypothesis:

$H_{0\text{-}5}$: The current mutagenicity of grapevine TE families cannot be linked to predictors of recent historical activity within the grapevine genome sequence data (see 1.6.5).

9.1.7 The generation of new clonal material by transposon mutagenesis

A population of 180 vines was produced from regenerated EC cultures. Among these plants both phenotypic variation and transposon polymorphisms, detected by S-SAP, were seen. Whole-genome sequencing performed on a subset of this population revealed the true degree of transposon variation: an average of 64 new insertions per individual, approximately 2 of which were in genic regions. These results show the potential of endogenous transposons for the generation of a mutation-saturated population.

These vines are the mitotic progeny of elite Pinot noir clone UCD5, and as such can be considered new clones of this variety. The lack of transgenic modification makes this source of material directly suitable for industry use. The association of particular phenotypes with genotypic change remains to be done. However, the population serves to disprove the following null hypothesis:

$H_{0\text{-}6}$: It is not possible to identify distinct phenotypic abnormalities in a population of grapevines regenerated after transposon mobilisation (see 1.6.6).

9.2 Future research

This project was initiated in order to discover whether endogenous transposons could be used to address the grapevine industry's need for increased genetic variation, as emphasised by Bonfiglioli and Hoskins (2008), without compromising the value of varietal identity. The mutant vines produced to date serve as a pilot population, demonstrating that this appears to be achievable. Phenotypic analysis of the regenerated vines in future seasons will address the degree of phenotypic variation among the vines and their similarity to the parental clone.

The wealth of information generated by producing and sequencing this population offers several immediate avenues of research.

9.2.1 Detailed analysis of TE insertion sites

Due to the time constraints of this project, the analysis of genomic sequence data of the regenerated vines has involved a broad overview of which elements are responsible for new mutations and the general context of their insertions. However, there are still many questions that this data can be used to address. Of particular interest is the determination of any sequence insertion bias for individual TE families. Past publications have reported TE insertion preferences for euchromatic or heterochromatic DNA, nested repeat regions, specific gene regions (such as UTRs) and sites linked to the original insertion locus (see 2.2.3).

Since the class I elements in particular exist as highly repetitive insertion families, it is important to characterise which elements within these families are capable of transposition. The genomic sequence data of the parental vine used and the EC cultures should be useful in this regard. Future work will involve matching new insertions in the regenerated vines to specific founder elements in the parental genome.

9.2.2 What have we missed?

In order to identify new transposon insertions, the defined grapevine elements deposited in the Repbase Update database were used to inform the RetroSeq software searches. It is possible that certain new mutations may be associated with transposon families not previously defined. Several computational approaches have been published for identifying various transposable elements types in assembled sequence data (McCarthy and McDonald 2003; Darzentas *et al.* 2010; Han and Wessler 2010). Future work will involve using these to investigate whether uncharacterised transposons are partly responsible for mutagenesis in the vine population.

9.2.3 Functional genomics of mutant phenotypes

Several vines displaying interesting phenotypes were identified among the regenerated population. These vines have been passaged through winter dormancy and pruned to a consistent size. Future research will perform more rigorous phenotypic characterisation of forthcoming growth seasons, and attempt to associate phenotypic abnormalities with genotypic changes. A parallel reverse-genetics approach will define candidate genes of interest and screen the collection of transposon-mutant vines for novel alleles at these loci.

9.2.4 Generation of a saturation mutagenesis population for grapevine

Based on the success of this technique, our research group at Lincoln University is currently collaborating with Plant and Food Research (New Zealand) to produce a sufficient number of mutant vines to constitute a mutation-saturated population. The project is also currently being expanded to produce a similar mutant population of Sauvignon blanc. These populations will form a resource of international significance for plant functional genomic research, as well as to the grapevine industry.

9.3 Final comments

As the dust thrown up by the justified excitement of whole-genome sequencing projects begins to settle, it is becoming increasingly evident that there is more to the blueprint of life than a linear code. The surprisingly sparse gene regions are now known to be supplemented by the regulatory control of intergenic DNA, epigenetic DNA modifications and the dynamic activity of transposable elements. In a post-genomic era, new sequencing technologies and unprecedented computational power are facilitating research into how these factors govern the expression of DNA, thereby shaping phenotype from the cellular to the individual level. As our understanding in these areas increases, we can look forward to new biotechnological applications that will have substantial impact on the fields of medicine, forensics and agriculture.

Over the past half-century, we have progressed from our understanding of transposons as unstable loci in maize to ubiquitous drivers of genome evolution. The interaction between environmental change and transposon activity has been demonstrated in multiple species and the mutagenic capacity of particular active elements has been exploited in heterologous systems through transgenesis. The evidence reported here of multiple endogenous elements with mobilisation potential indicates not only that transgenesis may not always be necessary for transposon mutagenesis, but also that it may distract the researcher's attention from endogenous transposition events.

In the past single nucleotide polymorphisms (SNPs) have held centre stage in the search for mutant alleles. But in the case of these mutations only alterations to exons, which account for approximately 6.2% of the grapevine genome, are likely to have any phenotypic effect. Approximately one third of these are expected to be synonymous and those resulting in amino acid changes may still not affect protein function. In comparison, the large size of transposons renders all insertions into coding sequences liable to affect protein function. Examples of phenotypic change arising from promoter insertions have been discussed (see 2.2.10) and even insertions into introns may be disruptive enough to result in phenotypic change (LUEHRSEN AND WALBOT 1992).

The most compelling reason that the mutagenic role of transposable elements cannot be ignored is that individual elements are estimated to generate mutations at a rate between 10^4 and 10^5 times higher than the rate of SNP accumulation (BIEMONT AND VIEIRA 2006). Unlike SNP mutations, insertion bias is a factor that needs to be considered with regards to transposon mutagenesis. However, most elements that show any bias appear to preferentially insert in euchromatin, presumably because of the accessible nature of the DNA at these regions.

The challenge of crop improvement in long-lived perennial crops is complex and takes time to address, particularly in grapevine where issues of varietal identity are of prime importance to the market. Nevertheless, the unusually long history of vegetative propagation in this species offers a unique insight into the biological processes driving somatic mutation and how these may result in novel phenotypes. This research has shown that endogenous transposons can be used as a source of novel genetic and phenotypic variation, an approach that is of particular importance to vegetatively propagated crops in which new traits do not need to be stabilised by crosses (PETRI AND BURGOS 2005).

References

Alleweldt, G., and Possingham, J. (1988) Progress in grapevine breeding. *Theoretical and Applied Genetics,* 75**:** 669-673.

Alonso, J. M., Stepanova, A. N., Leisse, T. J., Kim, C. J., Chen, H., Shinn, P., . . . Ecker, J. R. (2003) Genome-wide insertional mutagenesis of *Arabidopsis thaliana*. *Science,* 301**:** 653-657.

Andrews, S. (2010) FASTQC: A quality control tool for high throughput sequence data. http://www.bioinformatics.babraham.ac.uk/projects/fastqc.

Angel, A., Song, J., Dean, C. and Howard, M. (2011) A Polycomb-based switch underlying quantitative epigenetic memory. *Nature,* 476**:** 105-108.

Arabidopsis Genome Initiative (2000) Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. *Nature,* 408**:** 796-815.

Arias-Bolzmann, L., Sak, O., Musalem, A., Lodish, L., Báez, R. and De Sousa, L. J. (2003) Wine pricing: The influence of country of origin, variety, and wine magazine ratings. *International Journal of Wine Marketing,* 15**:** 47-57.

Arroyo-García, R., Lefort, F., de Andrés, M. T., Ibáñez, J., Borrego, J., Jouve, N., . . . Martínez-Zapater, J. M. (2002) Chloroplast microsatellite polymorphisms in *Vitis* species. *Genome,* 45**:** 1142-1149.

Axtell, M. J., Jan, C., Rajagopalan, R. and Bartel, D. P. (2006) A two-hit trigger for siRNA biogenesis in plants. *Cell,* 127**:** 565-577.

Bairu, M. W., Aremu, A. O. and Van Staden, J. (2011) Somaclonal variation in plants: causes and detection methods. *Plant Growth Regulation,* 63**:** 147-173.

Baránek, M., Křižan, B., Ondrušíková, E. and Pidra, M. (2010) DNA-methylation changes in grapevine somaclones following in vitro culture and thermotherapy. *Plant Cell, Tissue and Organ Culture,* 101**:** 11-22.

Baribault, T., Skene, K., Cain, P. and Scott, N. S. (1990) Transgenic grapevines: Regeneration of shoots expressing β-glucuronidase. *Journal of Experimental Botany,* 41**:** 1045-1049.

Barnard, H., Dooley, A. N., Areshian, G., Gasparyan, B. and Faull, K. F. (2011) Chemical evidence for wine production around 4000 BCE in the Late Chalcolithic Near Eastern highlands. *Journal of Archaeological Science,* 38**:** 977-984.

Barret, P., Brinkman, M. and Beckert, M. (2006) A sequence related to rice Pong transposable element displays transcriptional activation by in vitro culture and reveals somaclonal variations in maize. *Genome,* 49**:** 1399-1407.

Bartee, L., Malagnac, F. and Bender, J. (2001) *Arabidopsis* cmt3 chromomethylase mutations block non-CG methylation and silencing of an endogenous gene. *Genes & Development,* 15**:** 1753-1758.

Barton, N. H., and Charlesworth, B. (1998) Why sex and recombination? *Science,* 281**:** 1986-1990.

Baulcombe, D. (2004) RNA silencing in plants. *Nature,* 431**:** 356-363.

Ben Amar, A., Cobanov, P., Boonrod, K., Krczal, G., Bouzid, S., Ghorbel, A. and Reustle, G. (2007) Efficient procedure for grapevine embryogenic suspension establishment and plant regeneration: role of conditioned medium for cell proliferation. *Plant Cell Reports,* 26**:** 1439-1447.

Benham, J., Jeung, J.-U., Jasieniuk, M., Kanazin, V. and Blake, T. (1999) Genographer: a graphical tool for automated fluorescent AFLP and microsatellite analysis. *Journal of Agricultural Genomics,* 4.

Benjak, A., Boue, S., Forneck, A. and Casacuberta, J. M. (2009) Recent amplification and impact of MITEs on the genome of grapevine (*Vitis vinifera* L.). *Genome Biology and Evolution,* 1**:** 75-84.

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

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

Bennetzen, J. L. (2005) Transposable elements, gene creation and genome rearrangement in flowering plants. *Current Opinion in Genetics & Development,* 15**:** 621-627.

Berenyi, M., Gichuki, S. T., Schmidt, J. and Burg, K. (2002) Ty1-copia retrotransposon-based S-SAP (sequence-specific amplified polymorphism) for genetic analysis of sweetpotato. *Theoretical and Applied Genetics,* 105**:** 862–869.

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

Bertsch, C., Kieffer, F., Maillot, P., Farine, S., Butterlin, G., Merdinoglu, D. and Walter, B. (2005) Genetic chimerism of *Vitis vinifera* cv. Chardonnay 96 is maintained through organogenesis but not somatic embryogenesis. *BMC Plant Biology,* 5**:** 20.

Biemont, C., and Vieira, C. (2006) Genetics: Junk DNA as an evolutionary force. *Nature,* 443**:** 521-524.

Boeke, J., and Corces, V. G. (1989) Transcription and reverse transcription of retrotransposons. *Annual Reviews in Microbiology,* 43**:** 403-434.

Bonfiglioli, R., and Hoskins, N. (2008) Sauvignon Blanc selection in Sancerre. *The Australian & New Zealand Grapegrower & Winemaker,* 530**:** 91-93.

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.

Bowers, J. E., and Meredith, C. P. (1997) The parentage of a classic wine grape, Cabernet Sauvignon. *Nature Genetics,* 16**:** 84-87.

Bragato, R. (1895) *Report on the prospects of viticulture in New Zealand: Together with instructions for planting and pruning*. Government Printer.

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

Bray, C. M., and West, C. E. (2005) DNA repair mechanisms in plants: crucial sensors and effectors for the maintenance of genome integrity. *New Phytologist,* 168**:** 511-528.

Bretó, M., Ruiz, C., Pina, J. and Asíns, M. (2001) The diversification of *Citrus clementina* Hort. ex Tan., a vegetatively propagated crop species. *Molecular Phylogenetics and Evolution,* 21**:** 285-293.

Britt, A. B. (1999) Molecular genetics of DNA repair in higher plants. *Trends in Plant Science,* 4**:** 20-25.

Byun, S. O., Fang, Q., Zhou, H. and Hickford, J. G. H. (2009) An effective method for silver-staining DNA in large numbers of polyacrylamide gels. *Analytical Biochemistry,* 385**:** 174-175.

Campbell, B. C., LeMare, S., Piperidis, G. and Godwin, I. D. (2011) IRAP, a retrotransposon-based marker system for the detection of somaclonal variation in barley. *Molecular Breeding,* 27**:** 193-206.

Carcamo, C., Provedo, I. and Arroyo-Garcia, R. (2010) *Polymorphism in ancient Tempranillo grapevine clones detected using microsatellite and retrotransposon markers*.

Carimi, F., Barizza, E., Gardiman, M. and Schiavo, F. (2005) Somatic embryogenesis from stigmas and styles of grapevine. *In Vitro Cellular & Developmental Biology - Plant,* 41**:** 249-252.

Carpenter, R., and Coen, E. S. (1990) Floral homeotic mutations produced by transposon-mutagenesis in Antirrhinum majus. *Genes & Development,* 4**:** 1483-1493.

Carpenter, R., Martin, C. and Coen, E. S. (1987) Comparison of genetic behaviour of the transposable element Tam3 at two unlinked pigment loci in Antirrhinum majus. *Molecular and General Genetics MGG,* 207**:** 82-89.

Carrier, G., Le Cunff, L., Dereeper, A., Legrand, D., Sabot, F., Bouchez, O., . . . This, P. (2012) Transposable elements are a major cause of somatic polymorphism in *Vitis vinifera* L. *PloS one,* 7**:** e32973.

Castro, I., D'Onofrio, C., Martín, J. P., Ortiz, J. M., De Lorenzis, G., Ferreira, V. and Pinto-Carnide, O. (2012) Effectiveness of AFLPs and retrotransposon-based markers for the identification of portuguese grapevine cultivars and clones. *Molecular Biotechnology***:** 1-14.

Chaïb, J., Torregrosa, L., Mackenzie, D., Corena, P., Bouquet, A. and Thomas, M. R. (2010) The grape microvine - a model system for rapid forward and reverse genetics of grapevines. *The Plant Journal,* 62**:** 1083-1092.

Chalvet, F., Grimaldi, C., Kaper, F., Langin, T. and Daboussi, M.-J. (2003) Hop, an active Mutator-like element in the genome of the fungus Fusarium oxysporum. *Molecular biology and evolution,* 20**:** 1362-1375.

Chan, S. W.-L., Henderson, I. R. and Jacobsen, S. E. (2005) Gardening the genome: DNA methylation in *Arabidopsis thaliana*. *Nature Reviews Genetics,* 6**:** 351-360.

Chen, I., and Manchester, S. R. (2007) Seed morphology of modern and fossil Ampelocissus (Vitaceae) and implications for phytogeography. *American Journal of Botany,* 94**:** 1534-1553.

Choi, Y., Gehring, M., Johnson, L., Hannon, M., Harada, J. J., Goldberg, R. B., . . . Fischer, R. L. (2002) DEMETER, a DNA glycosylase domain protein, is required for endosperm gene imprinting and seed viability in *Arabidopsis*. *Cell,* 110**:** 33-42.

Coffey, A. J., Kokocinski, F., Calafato, M. S., Scott, C. E., Palta, P., Drury, E., . . . Hunt, S. (2011) The GENCODE exome: sequencing the complete human exome. *European Journal of Human Genetics,* 19**:** 827-831.

Cresse, A. D., Hulbert, S. H., Brown, W. E., Lucas, J. R. and Bennetzen, J. L. (1995) Mu1-related transposable elements of maize preferentially insert into low copy number DNA. *Genetics,* 140**:** 315-324.

D'Onofrio, C., De Lorenzis, G., Giordani, T., Natali, L., Scalabrelli, G. and Cavallini, A. (2009) Retrotransposon-based molecular markers in grapevine species and cultivars identification and phylogenetic analysis. *Acta Horticulturae,* 827**:** 45-52.

D'Onofrio, C., De Lorenzis, G., Giordani, T., Natali, L., Cavallini, A. and Scalabrelli, G. (2010) Retrotransposon-based molecular markers for grapevine species and cultivars identification. *Tree Genetics & Genomes,* 6**:** 451-466.

Darwin, C. (1870) *On the origin of species by means of natural selection*. D. Appleton and Company, New York.

Darzentas, N., Bousios, A., Apostolidou, V. and Tsaftaris, A. S. (2010) MASiVE: Mapping and Analysis of SireVirus Elements in plant genome sequences. *Bioinformatics,* 26**:** 2452-2454.

Das, D., Reddy, M., Upadhyaya, K. and Sopory, S. (2002) An efficient leaf-disc culture method for the regeneration via somatic embryogenesis and transformation of grape (*Vitis vinifera* L.). *Plant Cell Reports,* 20**:** 999-1005.

Dhekney, S. A., Li, Z. T., Compton, M. E. and Gray, D. J. (2009) Optimizing initiation and maintenance of *Vitis* embryogenic cultures. *HortScience,* 44**:** 1400-1406.

Dietrich, C. R., Cui, F., Packila, M. L., Li, J., Ashlock, D. A., Nikolau, B. J. and Schnable, P. S. (2002) Maize Mu transposons are targeted to the 5' untranslated region of the gl8 gene and sequences flanking Mu target-site duplications exhibit nonrandom nucleotide composition throughout the genome. *Genetics,* 160**:** 697-716.

Doligez, A., Bouquet, A., Danglot, Y., Lahogue, F., Riaz, S., Meredith, C., . . . This, P. (2002) Genetic mapping of grapevine (*Vitis vinifera* L.) applied to the detection of QTLs for seedlessness and berry weight. *Theoretical and Applied Genetics,* 105**:** 780-795.

Domingues, D. S., Cruz, G. M., Metcalfe, C. J., Nogueira, F. T., Vicentini, R., de S Alves, C. and Van Sluys, M.-A. (2012) Analysis of plant LTR-retrotransposons at the fine-scale family level reveals individual molecular patterns. *BMC Genomics,* 13**:** 137.

Dowen, R. H., Pelizzola, M., Schmitz, R. J., Lister, R., Dowen, J. M., Nery, J. R., . . . Ecker, J. R. (2012) Widespread dynamic DNA methylation in response to biotic stress. *Proceedings of the National Academy of Sciences,* 109**:** E2183-E2191.

Doyle, J. J., and Doyle, J. L. (1990) Isolation of plant DNA from fresh tissue. *Focus,* 12**:** 13-15.

Du, J., Tian, Z., Hans, C. S., Laten, H. M., Cannon, S. B., Jackson, S. A., . . . Ma, J. (2010) Evolutionary conservation, diversity and specificity of LTR-retrotransposons in flowering plants: insights from genome-wide analysis and multi-specific comparison. *The Plant Journal,* 63**:** 584-598.

Edgar, R. C. (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research,* 32**:** 1792-1797.

Ellis, T. H. N., Poyser, S. J., Knox, M. R., Vershinin, A. V. and Ambrose, M. J. (1998) Polymorphism of insertion sites of Ty1-copia class retrotransposons and its use for linkage and diversity analysis in pea. *Molecular and General Genetics MGG,* 260**:** 9-19.

Favory, J.-J., Stec, A., Gruber, H., Rizzini, L., Oravecz, A., Funk, M., . . . Oakeley, E. J. (2009) Interaction of COP1 and UVR8 regulates UV-B-induced photomorphogenesis and stress acclimation in *Arabidopsis*. *The EMBO journal,* 28**:** 591-601.

Felsenstein, J. (1974) The evolutionary advantage of recombination. *Genetics,* 78**:** 737-756.

Ferguson, A. A., and Jiang, N. (2011) Pack-MULEs: recycling and reshaping genes through GC-biased acquisition. *Mobile Genetic Elements,* 1**:** 135-138.

Fernandez, L., Chaïb, J., Martinez-Zapater, J.-M., Thomas, M. R. and Torregrosa, L. (2013) Mis-expression of a PISTILLATA-like MADS box gene prevents fruit development in grapevine. *The Plant Journal*.

Fernandez, L., Doligez, A., Lopez, G., Thomas, M. R., Bouquet, A. and Torregrosa, L. (2006) Somatic chimerism, genetic inheritance, and mapping of the fleshless berry (flb) mutation in grapevine (*Vitis vinifera* L.). *Genome,* 49**:** 721-728.

Fernandez, L., Torregrosa, L., Segura, V., Bouquet, A. and Martinez-Zapater, J. M. (2010) Transposon-induced gene activation as a mechanism generating cluster shape somatic variation in grapevine. *The Plant Journal,* 61**:** 545-557.

Feschotte, C., Jiang, N. and Wessler, S. R. (2002) Plant transposable elements: where genetics meets genomics. *Nature Reviews Genetics,* 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. *Molecular Biology and Evolution,* 17**:** 730-737.

Feschotte, C., and Pritham, E. J. (2007) DNA transposons and the evolution of eukaryotic genomes. *Annual Review of Genetics,* 41**:** 331.

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

Flavell, R. B. (1994) Inactivation of gene expression in plants as a consequence of specific sequence duplication. *Proceedings of the National Academy of Sciences,* 91**:** 3490-3496.

Fleet, G. H. (2008) Wine yeasts for the future. *FEMS Yeast Research,* 8**:** 979-995.

Fleige, S., and Pfaffl, M. W. (2006) RNA integrity and the effect on the real-time qRT-PCR performance. *Molecular Aspects of Medicine,* 27**:** 126-139.

Fournier-Level, A., Lacombe, T., Le Cunff, L., Boursiquot, J. M. and This, P. (2010) Evolution of the VvMybA gene family, the major determinant of berry colour in cultivated grapevine (*Vitis vinifera* L.). *Heredity,* 104**:** 351-362.

Francis, K. E., and Spiker, S. (2005) Identification of *Arabidopsis thaliana* transformants without selection reveals a high occurrence of silenced T-DNA integrations. *The Plant Journal,* 41**:** 464-477.

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

Franks, T., Gang He, D. and Thomas, M. (1998) Regeneration of transgenic *Vitis vinifera* L. Sultana plants: genotypic and phenotypic analysis. *Molecular Breeding,* 4**:** 321-333.

Fujino, K., Hashida, S.-n., Ogawa, T., Natsume, T., Uchiyama, T., Mikami, T. and Kishima, Y. (2011) Temperature controls nuclear import of Tam3 transposase in Antirrhinum. *The Plant Journal,* 65**:** 146-155.

Fujino, K., Sekiguchi, H. and Kiguchi, T. (2005) Identification of an active transposon in intact rice plants. *Molecular Genetics and Genomics,* 273**:** 150-157.

Fukai, E., Soyano, T., Umehara, Y., Nakayama, S., Hirakawa, H., Tabata, S., . . . Hayashi, M. (2012) Establishment of a Lotus japonicus gene tagging population using the exon-targeting endogenous retrotransposon LORE1. *The Plant Journal,* 69**:** 720-730.

Fukai, E., Umehara, Y., Sato, S., Endo, M., Kouchi, H., Hayashi, M., . . . Hirochika, H. (2010) Derepression of the plant chromovirus LORE1 induces germline transposition in regenerated plants. *PLoS genetics,* 6**:** e1000868.

Gabriel, A., Dapprich, J., Kunkel, M., Gresham, D., Pratt, S. C. and Dunham, M. J. (2006) Global mapping of transposon location. *PLoS Genetics,* 2**:** e212.

Gambino, G., Ruffa, P., Vallania, R. and Gribaudo, I. (2007) Somatic embryogenesis from whole flowers, anthers and ovaries of grapevine (*Vitis* spp.). *Plant Cell, Tissue and Organ Culture,* 90**:** 79-83.

Gao, D. (2012) Identification of an active Mutator-like element (MULE) in rice (Oryza sativa). *Molecular Genetics and Genomics,* 287**:** 261-271.

Gelvin, S. B. (2003) *Agrobacterium*-mediated plant transformation: the biology behind the "gene-jockeying" tool. *Microbiology and Molecular Biology Reviews,* 67**:** 16-37.

Gelvin, S. B. (2005) Agricultural biotechnology: Gene exchange by design. *Nature,* 433**:** 583-584.

Gendrel, A.-V., Lippman, Z., Yordan, C., Colot, V. and Martienssen, R. A. (2002) Dependence of heterochromatic histone H3 methylation patterns on the *Arabidopsis* gene DDM1. *Science,* 297**:** 1871-1873.

Girard, A., and Hannon, G. J. (2008) Conserved themes in small-RNA-mediated transposon control. *Trends in Cell Biology,* 18**:** 136-148.

Gleave, A. P. (1992) A versatile binary vector system with a T-DNA organisational structure conducive to efficient integration of cloned DNA into the plant genome. *Plant Molecular Biology,* 20**:** 1203-1207.

Goecks, J., Nekrutenko, A., Taylor, J. and Team, T. G. (2010) Galaxy: a comprehensive approach for supporting accessible, reproducible, and transparent computational research in the life sciences. *Genome Biology,* 11**:** R86.

Gognies, S., Belarbi, A. and Ait Barka, E. (2001) Saccharomyces cerevisiae, a potential pathogen towards grapevine, *Vitis vinifera*. *FEMS Microbiology Ecology,* 37**:** 143-150.

Graham, T., and Boissinot, S. (2006) The genomic distribution of L1 elements: the role of insertion bias and natural selection. *Journal of Biomedicine and Biotechnology,* 2006.

Grandbastien, M. (1998) Activation of plant retrotransposons under stress conditions. *Trends in Plant Science,* 3**:** 181-187.

Grassi, F., De Mattia, F., Zecca, G., Sala, F. and Labra, M. (2008) Historical isolation and Quaternary range expansion of divergent lineages in wild grapevine. *Biological Journal of the Linnean Society,* 95**:** 611-619.

Grassi, F., Labra, M., Imazio, S., Rubio, R. O., Failla, O., Scienza, A. and Sala, F. (2006) Phylogeographical structure and conservation genetics of wild grapevine. *Conservation Genetics,* 7**:** 837-845.

Grassi, F., Labra, M., Imazio, S., Spada, A., Sgorbati, S., Scienza, A. and Sala, F. (2003) Evidence of a secondary grapevine domestication centre detected by SSR analysis. *Theoretical and Applied Genetics,* 107**:** 1315-1320.

Grewal, S. I. (2010) RNAi-dependent formation of heterochromatin and its diverse functions. *Current Opinion in Genetics & Development,* 20**:** 134-141.

Han, Y., and Wessler, S. R. (2010) MITE-Hunter: a program for discovering miniature inverted-repeat transposable elements from genomic sequences. *Nucleic acids research,* 38**:** e199-e199.

Hansen, C., and Heslop-Harrison, J. S. (2004) Sequences and phylogenies of plant pararetroviruses, viruses, and transposable elements. *Advances in Botanical Research,* 41**:** 165-193.

Harrison, B. J., and Fincham, J. R. S. (1964) Instability at the Pal locus in *Antirrhinum majus*. *Heredity,* 19**:** 237-258.

Hasegawa, M., Kishino, H. and Yano, T.-a. (1985) Dating of the human-ape splitting by a molecular clock of mitochondrial DNA. *Journal of Molecular Evolution,* 22**:** 160-174.

Hensel, M., Shea, J. E., Gleeson, C., Jones, M. D., Dalton, E. and Holden, D. W. (1995) Simultaneous identification of bacterial virulence genes by negative selection. *Science,* 269**:** 400-403.

Herr, A., Jensen, M., Dalmay, T. and Baulcombe, D. (2005) RNA polymerase IV directs silencing of endogenous DNA. *Science,* 308**:** 118-120.

Heslop-Harrison, J. S., Brandes, A., Taketa, S., Schmidt, T., Vershinin, A., Alkhimova, E., . . . Harrison, G. (1997) The chromosomal distributions of Ty1-copia group retrotransposable elements in higher plants and their implications for genome evolution. *Genetica,* 100**:** 197-204.

Higo, K., Ugawa, Y., Iwamoto, M. and Korenaga, T. (1999) Plant cis-acting regulatory DNA elements (PLACE) database: 1999. *Nucleic Acids Research,* 27**:** 297-300.

Hirochika, H. (1997) Retrotransposons of rice: their regulation and use for genome analysis. *Plant Molecular Biology,* 35**:** 231-240.

Hirochika, H., Sugimoto, K., Otsuki, Y., Tsugawa, H. and Kanda, M. (1996) Retrotransposons of rice involved in mutations induced by tissue culture. *Proceedings of the National Academy of Sciences,* 93**:** 7783-7788.

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

Holden, M. (1950) A study of enzymes that can break down tobacco-leaf components; 5. Unfractionated fungal enzymes. *Biochemical Journal,* 47**:** 426-431.

Holmberg, A., Blomstergren, A., Nord, O., Lukacs, M., Lundeberg, J. and Uhlén, M. (2005) The biotin-streptavidin interaction can be reversibly broken using water at elevated temperatures. *Electrophoresis,* 26**:** 501-510.

Horsch, R. B., Fry, J. E., Hoffmann, N. L., Eichholtz, D., Rogers, S. G. a. and Fraley, R. T. (1985) A simple and general method for transferring genes into plants. *Science,* 227**:** 1229-1231.

Hřibová, E., Neumann, P., Matsumoto, T., Roux, N., Macas, J. and Doležel, J. (2010) Repetitive part of the banana (*Musa acuminata*) genome investigated by low-depth 454 sequencing. *BMC Plant Biology,* 10**:** 204.

Hsieh, T.-F., Ibarra, C. A., Silva, P., Zemach, A., Eshed-Williams, L., Fischer, R. L. and Zilberman, D. (2009) Genome-wide demethylation of *Arabidopsis* endosperm. *Science,* 324**:** 1451-1454.

Hu, W., Das, O. P. and Messing, J. (1995) Zeon-1, a member of a new maize retrotransposon family. *Molecular and General Genetics MGG,* 248**:** 471-480.

Huang, J., Zhang, K., Shen, Y., Huang, Z., Li, M., Tang, D., . . . Cheng, Z. (2009) Identification of a high frequency transposon induced by tissue culture, nDaiZ, a member of the hAT family in rice. *Genomics,* 93**:** 274-281.

Iida, S., Morita, Y., Choi, J.-D., Park, K.-I. and Hoshino, A. (2004) Genetics and epigenetics in flower pigmentation associated with transposable elements in morning glories. *Advances in biophysics,* 38**:** 141-159.

Iocco, P., Franks, T. and Thomas, M. R. (2001) Genetic transformation of major wine grape cultivars of *Vitis Vinifera* L. *Transgenic Research,* 10**:** 105-112.

Ito, H., Gaubert, H., Bucher, E., Mirouze, M., Vaillant, I. and Paszkowski, J. (2011) An siRNA pathway prevents transgenerational retrotransposition in plants subjected to stress. *Nature,* 472**:** 115-119.

Jaillon, O., Aury, J.-M., Noel, B., Policriti, A., Clepet, C., Casagrande, A., . . . Jubin, C. (2007) The grapevine genome sequence suggests ancestral hexaploidization in major angiosperm phyla. *Nature,* 449**:** 463-467.

Jander, G., Baerson, S. R., Hudak, J. A., Gonzalez, K. A., Gruys, K. J. and Last, R. L. (2003) Ethylmethanesulfonate saturation mutagenesis in *Arabidopsis* to determine frequency of herbicide resistance. *Plant Physiology,* 131**:** 139-146.

Janssen, B.-J., and Gardner, R. (1990) Localized transient expression of GUS in leaf discs following cocultivation with Agrobacterium. *Plant Molecular Biology,* 14**:** 61-72.

Jefferson, R. A., Kavanagh, T. A. and Bevan, M. W. (1987) GUS fusions: beta-glucuronidase as a sensitive and versatile gene fusion marker in higher plants. *EMBO J.,* 6**:** 3901–3907.

Jeffreys, A. J., Wilson, V. and Thein, S. L. (1985) Hypervariable, 'minisatellite' regions in human DNA. *Nature,* 314**:** 67-73.

Jeon, J.-S., Lee, S., Jung, K.-H., Jun, S.-H., Jeong, D.-H., Lee, J., . . . An, G. (2000) T-DNA insertional mutagenesis for functional genomics in rice. *The Plant Journal,* 22**:** 561-570.

Jiang, C., Mithani, A., Gan, X., Belfield, E. J., Klingler, J. P., Zhu, J.-K., . . . Harberd, N. P. (2011) Regenerant *Arabidopsis* lineages display a distinct genome-wide spectrum of mutations conferring variant phenotypes. *Current biology : CB,* 21**:** 1385-1390.

Jiang, H.-E., Zhang, Y.-B., Li, X., Yao, Y.-F., Ferguson, D. K., Lvº, E.-G. and Li, C.-S. (2009) Evidence for early viticulture in China: proof of a grapevine (*Vitis vinifera* L., Vitaceae) in the Yanghai Tombs, Xinjiang. *Journal of Archaeological Science,* 36**:** 1458-1465.

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, C., Edwards, K., Castaglione, S., Winfield, M., Sala, F., Van de Wiel, C., . . . Daly, A. (1997) Reproducibility testing of RAPD, AFLP and SSR markers in plants by a network of European laboratories. *Molecular Breeding,* 3**:** 381-390.

Jordan, B. R. (2002) Review: Molecular response of plant cells to UV-B stress. *Functional Plant Biology,* 29**:** 909-916.

Jordan, I. K., Rogozin, I. B., Glazko, G. V. and Koonin, E. V. (2003) Origin of a substantial fraction of human regulatory sequences from transposable elements. *Trends in Genetics,* 19**:** 68-72.

Jurka, J., Kapitonov, V. V., Pavlicek, A., Klonowski, P., Kohany, O. and Walichiewicz, J. (2005) Repbase Update, a database of eukaryotic repetitive elements. *Cytogenetic and Genome Research,* 110**:** 462-467.

Kalendar, R., Grob, T., Regina, M., Suoniemi, A. and Schulman, A. (1999) IRAP and REMAP: two new retrotransposon-based DNA fingerprinting techniques. *Theoretical and Applied Genetics,* 98**:** 704-711.

Kalendar, R., and Schulman, A. H. (2007) IRAP and REMAP for retrotransposon-based genotyping and fingerprinting. *Nature Protocols,* 1**:** 2478-2484.

Kapitonov, V. V., and Jurka, J. (2007) *Helitrons* on a roll: eukaryotic rolling-circle transposons. *TRENDS in Genetics,* 23**:** 521-529.

Karimi, M., Inzé, D. and Depicker, A. (2002) GATEWAY vectors for *Agrobacterium*-mediated plant transformation. *Trends in Plant Science,* 7**:** 193-195.

Kazazian, H. H. (2004) Mobile elements: drivers of genome evolution. *Science,* 303**:** 1626-1632.

Keane, T. M., Wong, K. and Adams, D. J. (2013) RetroSeq: transposable element discovery from next-generation sequencing data. *Bioinformatics,* 29**:** 389-390.

Kent, W. J. (2002) BLAT—the BLAST-like alignment tool. *Genome Research,* 12**:** 656-664.

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

Kikkert, J., Striem, M., Vidal, J., Wallace, P., Barnard, J. and Reisch, B. (2005) Long-term study of somatic embryogenesis from anthers and ovaries of 12 grapevine (*Vitis* sp.) genotypes. *In Vitro Cellular and Development Biology - Plant,* 41**:** 232-239.

Kikkert, J. R., Striem, M. J., Martens, M. H., Wallace, P. G. and Reisch, B. I. (1997) Somatic embryogenesis from anthers and ovaries of six grapevine (*Vitis* sp.) cultivars. *In Vitro Cell Dev Biol Plant,* 33**:** 314-315.

Kikuchi, K., Terauchi, K., Wada, M. and Hirano, H.-Y. (2003) The plant MITE mPing is mobilized in anther culture. *Nature,* 421**:** 167-170.

Kim, D.-H., Doyle, M. R., Sung, S. and Amasino, R. M. (2009) Vernalization: winter and the timing of flowering in plants. *Annual Review of Cell and Developmental,* 25**:** 277-299.

Knox, M., Moreau, C., Lipscombe, J., Baker, D. and Ellis, N. (2009) High-throughput retrotransposon-based fluorescent markers: improved information content and allele discrimination. *Plant Methods,* 5**:** 10.

Kobayashi, S., Goto-Yamamoto, N. and Hirochika, H. (2005) Association of *VvmybA1* gene expression with anthocyanin production in grape. *Journal of the Japanese Society for Horticultural Science,* 74**:** 196-203.

Kobayashi, S., Ishimaru, M., Ding, C. K., Yakushiji, H. and Goto, N. (2004) Retrotransposon-induced mutations in grape skin colour. *Science,* 304**:** 982.

Kolesnik, T., Szeverenyi, I., Bachmann, D., Kumar, C. S., Jiang, S., Ramamoorthy, R., . . . Ramachandran, S. (2004) Establishing an efficient Ac/Ds tagging system in rice: large-scale analysis of Ds flanking sequences. *The Plant Journal,* 37**:** 301-314.

Komatsu, M., Shimamoto, K. and Kyozuka, J. (2003) Two-step regulation and continuous retrotransposition of the rice LINE-type retrotransposon Karma. *The Plant Cell Online,* 15**:** 1934-1944.

Koornneeff, M., Dellaert, L. W. M. and Van der Veen, J. H. (1982) EMS-and relation-induced mutation frequencies at individual loci in *Arabidopsis thaliana* (L.) Heynh. *Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis,* 93**:** 109-123.

Krzywinski, M. I., Schein, J. E., Birol, I., Connors, J., Gascoyne, R., Horsman, D., . . . Marra, M. A. (2009) Circos: An information aesthetic for comparative genomics. *Genome Research*.

Kuksova, V., Piven, N. and Gleba, Y. (1997) Somaclonal variation and in vitro induced mutagenesis in grapevine. *Plant Cell, Tissue and Organ Culture,* 49**:** 17-27.

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

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

Kunze, R., Saedler, H. and Lönnig, W. E. (1997) Plant Transposable Elements, pp. 331-470 in *Advances in Botanical Research*, edited by J. A. Callow. Academic Press.

Kwok, P.-Y. (2001) Methods for genotyping single nucleotide polymorphisms. *Annual review of genomics and human genetics,* 2**:** 235-258.

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

Langmead, B., and Salzberg, S. L. (2012) Fast gapped-read alignment with Bowtie 2. *Nature methods,* 9**:** 357-359.

Larkin, P. J., and Scowcroft, W. R. (1981) Somaclonal variation — a novel source of variability from cell cultures for plant improvement. *Theoretical and Applied Genetics,* 60**:** 197-214.

Law, J. A., and Jacobsen, S. E. (2010) Establishing, maintaining and modifying DNA methylation patterns in plants and animals. *Nature Reviews Genetics,* 11**:** 204-220.

Le, Q., Melayah, D., Bonnivard, E., Petit, M. and Grandbastien, M.-A. (2007) Distribution dynamics of the Tnt1 retrotransposon in tobacco. *Molecular Genetics and Genomics,* 278**:** 639-651.

Lemaire, O., Moneyron, A., Masson, J. E. and The Local Monitoring, C. (2010) "Interactive technology assessment" and beyond: the field trial of genetically modified grapevines at INRA-Colmar. *PLoS Biol,* 8**:** e1000551.

Lerat, E. (2009) Identifying repeats and transposable elements in sequenced genomes: how to find your way through the dense forest of programs. *Heredity,* 104**:** 520-533.

Li, R., Ye, J., Li, S., Wang, J., Han, Y., Ye, C., . . . Wong, G. K.-S. (2005) ReAS: Recovery of ancestral sequences for transposable elements from the unassembled reads of a whole genome shotgun. *PLoS Computational Biology,* 1**:** e43.

Li, X., Song, Y., Century, K., Straight, S., Ronald, P., Dong, X., . . . Zhang, Y. (2001) A fast neutron deletion mutagenesis-based reverse genetics system for plants. *The Plant Journal,* 27**:** 235-242.

Li, Z., Dhekney, S. and Gray, D. (2011) Use of the VvMybA1 gene for non-destructive quantification of promoter activity via color histogram analysis in grapevine (*Vitis vinifera*) and tobacco. *Transgenic Research***:** 1-11.

Lijavetzky, D., Ruiz-García, L., Cabezas, J. A., De Andrés, M. a. T., Bravo, G., Ibáñez, A., . . . Martínez-Zapater, J. M. (2006) Molecular genetics of berry colour variation in table grape. *Molecular Genetics and Genomics,* 276**:** 427-435.

Lingner, J., Hughes, T. R., Shevchenko, A., Mann, M., Lundblad, V. and Cech, T. R. (1997) Reverse transcriptase motifs in the catalytic subunit of telomerase. *Science,* 276**:** 561-567.

Lippman, Z., Gendrel, A.-V., Black, M., Vaughn, M. W., Dedhia, N., Richard McCombie, W., . . . Martienssen, R. (2004) Role of transposable elements in heterochromatin and epigenetic control. *Nature,* 430**:** 471-476.

Lippman, Z., and Martienssen, R. (2004) The role of RNA interference in heterochromatic silencing. *Nature,* 431**:** 364-370.

Lisch, D. (2002) Mutator transposons. *Trends in Plant Science,* 7**:** 498-504.

Lisch, D., and Bennetzen, J. L. (2011) Transposable element origins of epigenetic gene regulation. *Current Opinion in Plant Biology,* 14**:** 156-161.

Lisch, D., Chomet, P. and Freeling, M. (1995) Genetic characterization of the Mutator system in maize: behavior and regulation of Mu transposons in a minimal line. *Genetics,* 139**:** 1777.

Liu, S., Yeh, C.-T., Ji, T., Ying, K., Wu, H., Tang, H. M., . . . Schnable, P. S. (2009) Mu transposon insertion sites and meiotic recombination events co-localize with epigenetic marks for open chromatin across the maize genome. *PLoS Genetics,* 5**:** e1000733.

Liu, X., Yu, C. W., Duan, J., Luo, M., Wang, K., Tian, G., . . . Wu, K. (2011) HDA6 directly interacts with DNA methyltransferase MET1 and maintains transposable elements silencing in *Arabidopsis*. *Plant Physiology*.

Llorens, C., Futami, R., Covelli, L., Domínguez-Escribá, L., Viu, J. M., Tamarit, D., . . . Bernet, G. P. (2011) The Gypsy Database (GyDB) of mobile genetic elements: release 2.0. *Nucleic Acids Research,* 39**:** D70-D74.

Llorens, C., Muñoz-Pomer, A., Bernad, L., Botella, H. and Moya, A. (2009) Network dynamics of eukaryotic LTR retroelements beyond phylogenetic trees. *Biol Direct,* 4**:** 41.

Luehrsen, K. R., and Walbot, V. (1992) Insertion of non-intron sequence into maize introns interferes with splicing. *Nucleic acids research,* 20**:** 5181-5187.

Malacarne, G., Perazzolli, M., Cestaro, A., Sterck, L., Fontana, P., Van de Peer, Y., . . . Salamini, F. (2012) Deconstruction of the (paleo) polyploid grapevine genome based on the analysis of transposition events involving NBS resistance genes. *PLoS ONE,* 7**:** e29762.

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, pp. 281-295 in *Plant Gene Silencing*, edited by M. A. Matzke and A. J. M. Matzke. Springer Netherlands.

Mauro, M. C., Toutain, S., Walter, B., Pinck, L., Otten, L., Coutos-Thevenot, P., . . . Barbier, P. (1995) High efficiency regeneration of grapevine plants transformed with the GFLV coat protein gene. *Plant Science,* 112**:** 97-106.

May, B. P., Lippman, Z. B., Fang, Y., Spector, D. L. and Martienssen, R. A. (2005) Differential regulation of strand-specific transcripts from *Arabidopsis* centromeric satellite repeats. *PLoS Genetics,* 1**:** e79.

McCallum, C. M., Comai, L., Greene, E. A. and Henikoff, S. (2000) Targeting Induced LocalLesions IN Genomes (TILLING) for plant functional genomics. *Plant Physiology,* 123**:** 439-442.

McCarthy, E. M., and McDonald, J. F. (2003) LTR_STRUC: a novel search and identification program for LTR retrotransposons. *Bioinformatics,* 19**:** 362-367.

McClintock, B. (1948) Mutable Loci in Maize. *Carnegie Institute of Washington Year Book,* 47**:** 155-169.

McClintock, B. (1950) The origin and behavior of mutable loci in maize. *Proceedings of the National Academy of Sciences,* 36**:** 344-355.

McClintock, B. (1951) Mutable Loci in Maize. *Carnegie Institute of Washington Year Book,* 50**:** 174-181.

McCutcheon, E., Bruwer, J. and Li, E. (2009) Region of origin and its importance among choice factors in the wine-buying decision making of consumers. *International Journal of Wine Business Research,* 21**:** 212-234.

McGovern, P. E. (2003) *Ancient wine: the search for the origins of viniculture*. Princeton University Press, Princeton.

McGovern, P. E., Glusker, D. L., Exner, L. J. and Voigt, M. M. (1996) Neolithic resinated wine. *Nature,* 381**:** 480-481.

McGovern, P. E., Mirzoian, A. and Hall, G. R. (2009) Ancient Egyptian herbal wines. *Proceedings of the National Academy of Sciences,* 106**:** 7361-7366.

Melayah, D., Bonnivard, E., Chalhoub, B., Audeon, C. and Grandbastien, M.-A. (2001) The mobility of the tobacco Tnt1 retrotransposon correlates with its transcriptional activation by fungal factors. *The Plant Journal,* 28**:** 159-168.

Meyer, P. (2011) DNA methylation systems and targets in plants. *FEBS Letters,* 585**:** 2008-2015.

Mezzetti, B., Pandolfini, T., Navacchi, O. and Landi, L. (2002) Genetic transformation of *Vitis vinifera* via organogenesis. *BMC Biotechnology,* 2**:** 18.

Mhiri, C., Morel, J., Vernhettes, S., Casacuberta, J. M., H., L. and Grandbastien, M. (1997) The promoter of the tobacco Tnt1 retrotransposon is induced by wounding and by abiotic stress. *Plant Molecular Biology,* 33**:** 257-266.

Michaels, S. D., and Amasino, R. M. (1999) FLOWERING LOCUS C encodes a novel MADS domain protein that acts as a repressor of flowering. *The Plant Cell Online,* 11**:** 949-956.

Minami, T., Tanaka, T., Takasaki, S., Kawamura, K. and Hiratsuka, K. (2011) *In vivo* bioluminescence monitoring of defense gene expression in response to treatment with yeast cell wall extract. *Plant Biotechnology,* 28**:** 481-484.

Miura, A., Yonebayashi, S., Watanabe, K., Toyama, T., Shimada, H. and Kakutani, T. (2001) Mobilization of transposons by a mutation abolishing full DNA methylation in *Arabidopsis*. *Nature,* 411**:** 212-214.

Miyao, A., Nakagome, M., Ohnuma, T., Yamagata, H., Kanamori, H., Katayose, Y., . . . Hirochika, H. (2012) Molecular spectrum of somaclonal variation in regenerated rice revealed by whole-genome sequencing. *Plant and Cell Physiology,* 53**:** 256-264.

Moisy, C., Garrison, K., Meredith, C. and Pelsy, F. (2008) Characterization of ten novel Ty1/copia-like retrotransposon families of the grapevine genome. *BMC Genomics,* 9**:** 469.

Moon, S., Jung, K.-H., Lee, D.-e., Jiang, W.-Z., Koh, H. J., Heu, M.-H., . . . An, G. (2006) Identification of active transposon dTok, a member of the hAT family, in rice. *Plant and cell physiology,* 47**:** 1473-1483.

Moose, S. P., and Mumm, R. H. (2008) Molecular plant breeding as the foundation for 21st century crop improvement. *Plant Physiology,* 147**:** 969-977.

Muhammad, A., and Othman, F. (2005) Characterization of fusarium wilt-resistant and fusarium wilt-susceptible somaclones of banana cultivar rastali (Musa AAB) by random amplified polymorphic DNA and retrotransposon markers. *Plant Molecular Biology Reporter,* 23**:** 241-249.

Mullins, M., and Srinivasan, C. (1976) Somatic embryos and plantlets from an ancient clone of the grapevine (cv. Cabernet-Sauvignon) by apomixis in vitro. *Journal of Experimental Botany,* 27**:** 1022-1030.

Murashige, T., and Skoog, F. (1962) A revised medium for rapid growth and bio assays with tobacco tissue cultures. *Physiologia Plantarum,* 15**:** 473-497.

Myles, S., Boyko, A. R., Owens, C. L., Brown, P. J., Grassi, F., Aradhya, M. K., . . . Ware, D. (2011) Genetic structure and domestication history of the grape. *Proceedings of the National Academy of Sciences,* 108**:** 3530-3535.

Nacken, W. F., Piotrowiak, R., Saedler, H. and Sommer, H. (1991) The transposable element Tam1 from Antirrhinum majus shows structural homology to the maize transposon En/Spm and has no sequence specificity of insertion. *Molecular and General Genetics MGG,* 228**:** 201-208.

Naito, K., Zhang, F., Tsukiyama, T., Saito, H., Hancock, C. N., Richardson, A. O., . . . Wessler, S. R. (2009) Unexpected consequences of a sudden and massive transposon amplification on rice gene expression. *Nature,* 461**:** 1130-1134.

Nitsch, J., and Nitsch, C. (1969) Haploid plants from pollen grains. *Science,* 163**:** 85-87.

Nuthikattu, S., McCue, A. D., Panda, K., Fultz, D., DeFraia, C., Thomas, E. N. and Slotkin, R. K. (2013) The initiation of epigenetic silencing of active transposable elements is triggered by RDR6 and 21-22 nucleotide small interfering RNAs. *Plant Physiology*.

Ocaña, J., Walter, B. and Schellenbaum, P. (2013) Stable MSAP Markers for the Distinction of *Vitis vinifera* cv Pinot Noir Clones. *Molecular Biotechnology***:** 1-13.

Ohno, S. (1972) So much "junk" DNA in our genome. *Brookhaven symposia in biology,* 23**:** 366-370.

Olah, R., Zok, A., Pedryc, A., Howard, S. and Kovacs, L. G. (2009) Somatic embryogenesis in a broad spectrum of grape genotypes. *Scientia Horticulturae,* 120**:** 134-137.

Orgel, L. E., and Crick, F. H. (1980) Selfish DNA: the ultimate parasite. *Nature,* 284**:** 604.

Ovchinnikov, I., Troxel, A. B. and Swergold, G. D. (2001) Genomic characterization of recent human LINE-1 Insertions: evidence supporting random insertion. *Genome Research,* 11**:** 2050-2058.

Paran, I., and Michelmore, R. (1993) Development of reliable PCR-based markers linked to downy mildew resistance genes in lettuce. *Theoretical and Applied Genetics,* 85**:** 985-993.

Pardue, M.-L., and DeBaryshe, P. G. (2011) Retrotransposons that maintain chromosome ends. *Proceedings of the National Academy of Sciences,* 108**:** 20317-20324.

Pardue, M.-L., Rashkova, S., Casacuberta, E., DeBaryshe, P., George, J. and Traverse, K. (2005) Two retrotransposons maintain telomeres in Drosophila. *Chromosome Research,* 13**:** 443-453.

Pearce, S. R., Stuart-Rogers, C., Knox, M. R., Kumar, A., Ellis, T. H. N. and Flavell, A. J. (1999) Rapid isolation of plant Ty1-copia group retrotransposon LTR sequences for molecular marker studies. *The Plant Journal,* 19**:** 711-717.

Pecinka, A., Dinh, H. Q., Baubec, T., Rosa, M., Lettner, N. and Scheid, O. M. (2010) Epigenetic regulation of repetitive elements is attenuated by prolonged heat stress in *Arabidopsis*. *The Plant Cell Online,* 22**:** 3118-3129.

Pelsy, F. (2010) Molecular and cellular mechanisms of diversity within grapevine varieties. *Heredity,* 104**:** 331-340.

Pelsy, F., Hocquigny, S., Moncada, X., Barbeau, G., Forget, D., Hinrichsen, P. and Merdinoglu, D. (2010) An extensive study of the genetic diversity within seven French wine grape variety collections. *Theoretical and Applied Genetics,* 120**:** 1219-1231.

Pelsy, F., and 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.

Pelsy, F., Schehrer, L. and Merdinoglu, D. (2003) Development of grapevine retrotransposon-based molecular markers (S-SAP). *Acta Horticulturae,* 1**:** 83-88.

Pereira, H., Barão, A., Delgado, M., Morais-Cecílio, L. and Viegas, W. (2005) Genomic analysis of Grapevine Retrotransposon 1 (Gret1) in *Vitis vinifera*. *Theoretical and Applied Genetics,* 111**:** 871-878.

Pereira, V. (2004) Insertion bias and purifying selection of retrotransposons in the *Arabidopsis thaliana* genome. *Genome Biology,* 5**:** R79.

Perrin, M., Gertz, C. and Masson, J. E. (2004) High efficiency initiation of regenerable embryogenic callus from anther filaments of 19-grapevine genotypes grown worldwide. *Plant Science,* 167**:** 1343-1349.

Perrin, M., Martin, D., Joly, D., Demangeat, G., This, P. and Masson, J. E. (2001) Medium-dependent response of grapevine somatic embryogenic cells. *Plant Science,* 161**:** 107-116.

Peschke, V. M., and Phillips, R. L. (1991) Activation of the maize transposable element Suppressor-mutator (Spm) in tissue culture. *Theoretical and Applied Genetics,* 81**:** 90-97.

Peschke, V. M., Phillips, R. L. and Gengenbach, B. G. (1987) Discovery of transposable element activity among progeny of tissue culture—derived maize plants. *Science,* 238**:** 804-807.

Peterson-Burch, B. D., Wright, D. A., Laten, H. M. and Voytas, D. F. (2000) Retroviruses in plants? *Trends in Genetics,* 16**:** 151-152.

Petri, C., and Burgos, L. (2005) Transformation of fruit trees. Useful breeding tool or continued future prospect? *Transgenic Research,* 14**:** 15-26.

Pfaffl, M. W. (2006) *Relative quantification*. Taylor and Francis Group, New York.

Pouteau, S., Grandbastien, M.-A. and Boccara, M. (1994) Microbial elicitors of plant defence responses activate transcription of a retrotransposon. *The Plant Journal,* 5**:** 535-542.

Pouteau, S., Huttner, E., Grandbastien, M. and Caboche, M. (1991) Specific expression of the tobacco Tnt1 retrotransposon in protoplasts. *The EMBO journal,* 10**:** 1911.

Pritham, E. J., Putliwala, T. and Feschotte, C. (2007) *Mavericks*, a novel class of giant transposable elements widespread in eukaryotes and related to DNA viruses. *Gene,* 390**:** 3-17.

Rakocevic, A., Mondy, S., Tirichine, L., Cosson, V., Brocard, L., Iantcheva, A., . . . Ratet, P. (2009) MERE1, a low-copy-number copia-type retroelement in Medicago truncatula active during tissue culture. *Plant Physiol.,* 151**:** 1250-1263.

Raspor, P., Milek, D. M., Polanc, J., Možina, S. S. and Čadež, N. a. (2006) Yeasts isolated from three varieties of grapes cultivated in different locations of the Dolenjska vine-growing region, Slovenia. *International Journal of Food Microbiology,* 109**:** 97-102.

Ravo, M., Mutarelli, M., Ferraro, L., Grober, O. M. V., Paris, O., Tarallo, R., . . . Nola, E. (2008) Quantitative expression profiling of highly degraded RNA from formalin-fixed, paraffin-embedded breast tumor biopsies by oligonucleotide microarrays. *Laboratory Investigation,* 88**:** 430-440.

Regner, F., Stadlbauer, A. and Eisenheld, C. (2000) Molecular markers for genotyping grapevine and for identifying clones of traditional varieties, pp. 331-341 in *International Symposium on Molecular Markers for Characterizing Genotypes and Identifying Cultivars in Horticulture 546*.

Reid, K. E., Olsson, N., Schlosser, J., Peng, F. and Lund, S. T. (2006) An optimized grapevine RNA isolation procedure and statistical determination of reference genes for real-time RT-PCR during berry development. *BMC Plant Biology,* 6**:** 27.

Renouf, V., Miot-Sertier, C. c., Strehaiano, P. and Lonvaud-Funel, A. (2006) The wine microbial consortium: a real terrior characteristic. *International Journal of Vine and Wine Sciences,* 40**:** 209.

Riaz, S., Garrison, K. E., Dangl, G. S., Boursiquot, J.-M. and Meredith, C. P. (2002) Genetic divergence and chimerism within ancient asexually propagated winegrape cultivars. *Journal of the American Society for Horticultural Science,* 127**:** 508-514.

Rice, W. R. (2002) Experimental tests of the adaptive significance of sexual recombination. *Nature Reviews Genetics,* 3**:** 241-251.

Roudier, F., Ahmed, I., Bérard, C., Sarazin, A., Mary-Huard, T., Cortijo, S., . . . Al-Shikhley, L. (2011) Integrative epigenomic mapping defines four main chromatin states in *Arabidopsis*. *The EMBO journal,* 30**:** 1928-1938.

Roulin, A., Piegu, B., Fortune, P., Sabot, F. ß., D'Hont, A. l., Manicacci, D. and Panaud, O. (2009) Whole genome surveys of rice, maize and sorghum reveal multiple horizontal transfers of the LTR-retrotransposon Route66 in Poaceae. *BMC Evolutionary Biology,* 9**:** 58.

Sabot, F., and Schulman, A. (2006) Parasitism and the retrotransposon life cycle in plants: a hitchhiker's guide to the genome. *Heredity,* 97**:** 381-388.

Saenz, H. L., and Dehio, C. (2005) Signature-tagged mutagenesis: technical advances in a negative selection method for virulence gene identification. *Current Opinion in Microbiology,* 8**:** 612-619.

Sambrook, J., and Russell, D. W. (2001) *Molecular Cloning: A Laboratory Manual*. Cold Spring Harbor Laboratory Press.

SanMiguel, P., Tikhonov, A., Jin, Y.-K., Motchoulskaia, N., Zakharov, D., Melake-Berhan, A., . . . Avramova, Z. (1996) Nested retrotransposons in the intergenic regions of the maize genome. *Science,* 274**:** 765-768.

SantaLucia, J. (1998) A unified view of polymer, dumbbell, and oligonucleotide DNA nearest-neighbor thermodynamics. *Proceedings of the National Academy of Sciences,* 95**:** 1460-1465.

Santos-Rosa, M., Poutaraud, A., Merdinoglu, D. and Mestre, P. (2008) Development of a transient expression system in grapevine via agro-infiltration. *Plant Cell Reports,* 27**:** 1053-1063.

Scalabrin, S., Troggio, M., Moroldo, M., Pindo, M., Felice, N., Coppola, G., . . . Faes, G. (2010) Physical mapping in highly heterozygous genomes: a physical contig map of the Pinot Noir grapevine cultivar. *BMC Genomics,* 11**:** 204.

Schamel, G., and Anderson, K. (2003) Wine quality and varietal, regional and winery reputations: Hedonic prices for Australia and New Zealand. *Economic Record,* 79**:** 357-369.

Schellenbaum, P., Mohler, V., Wenzel, G. and Walter, B. (2008) Variation in DNA methylation patterns of grapevine somaclones (*Vitis vinifera* L.). *BMC plant biology,* 8**:** 78.

Schmid, C. W. (1998) Does SINE evolution preclude Alu function? *Nucleic Acids Research,* 26**:** 4541-4550.

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

Schnable, P. S., Ware, D., Fulton, R. S., Stein, J. C., Wei, F., Pasternak, S., . . . Wilson, R. K. (2009) The B73 Maize Genome: Complexity, Diversity, and Dynamics. *Science,* 326**:** 1112-1115.

Schneider, S., Reustle, G. and Zyprian, E. (1996) Detection of somaclonal variation in grapevine regenerants from protoplasts by RAPD-PCR. *Vitis,* 35**:** 99-100.

Schoor, O., Weinschenk, T., Hennenlotter, J., Corvin, S., Stenzl, A., Rammensee, H. G. and Stevanoviƒá, S. (2003) Moderate degradation does not preclude microarray analysis of small amounts of RNA. *BioTechniques,* 35**:** 1192-1196, 1198-1201.

Shahmuradov, I. A., Solovyev, V. V. and Gammerman, A. (2005) Plant promoter prediction with confidence estimation. *Nucleic Acids Research,* 33**:** 1069-1076.

Singer, T., Yordan, C. and Martienssen, R. A. (2001) Robertson's Mutator transposons in *A. thaliana* are regulated by the chromatin-remodeling gene Decrease in DNA Methylation (DDM1). *Genes & development,* 15**:** 591-602.

Slinkard, K. W., and Singleton, V. L. (1984) Phenol content of grape skins and the loss of ability to make anthocyanins by mutation. *Vitis,* 23**:** 175-178.

Slotkin, R. K., and Martienssen, R. (2007) Transposable elements and the epigenetic regulation of the genome. *Nature Reviews Genetics,* 8**:** 272-285.

Slotkin, R. K., Vaughn, M., Borges, F., Tanurdžić, M., Becker, J. D., Feijó, J. A. and Martienssen, R. A. (2009) Epigenetic reprogramming and small RNA silencing of transposable elements in pollen. *Cell,* 136**:** 461-472.

Stajner, N., Jakse, J., Javornik, B., Masuelli, R. W. and Martinez, L. E. (2009) Highly variable AFLP and S-SAP markers for the identification of Malbec and Syrah clones. *Vitis,* 48**:** 145-150.

Suprasanna, P., Manjunatha, B. R., Patade, V. Y., Desai, N. S. and Bapat, V. A. (2010) Profiling of culture-induced variation in sugarcane plants regenerated via direct and indirect somatic embryogenesis by using transposon-insertion polymorphism. *Sugar Tech,* 12**:** 26-30.

Syed, N. H., and Flavell, A. J. (2007) Sequence-specific amplification polymorphisms (SSAPs): a multi-locus approach for analyzing transposon insertions. *Nature Protocols,* 1**:** 2746-2752.

Tadege, M., Wen, J., He, J., Tu, H., Kwak, Y., Eschstruth, A., . . . Mysore, K. S. (2008) Large-scale insertional mutagenesis using the *Tnt1* retrotransposon in the model legume Medicago truncatula. *The Plant Journal,* 54**:** 335–347.

Takeda, S., Sugimoto, K., Otsuki, H. and Hirochika, H. (1999) A 13-bp cis-regulatory element in the LTR promoter of the tobacco retrotransposon Tto1 is involved in responsiveness to tissue culture, wounding, methyl jasmonate and fungal elicitors. *The Plant Journal,* 18**:** 383-393.

Tam, S. M., Mhiri, C., Vogelaar, A., Kerkveld, M., Pearce, S. R. and Grandbastien, M.-A. (2005) Comparative analyses of genetic diversities within tomato and pepper collections detected by retrotransposon-based SSAP, AFLP and SSR. *Theoretical and Applied Genetics,* 110**:** 819-831.

Tanurdzic, M., Vaughn, M. W., Jiang, H., Lee, T.-J., Slotkin, R. K., Sosinski, B., . . . Martienssen, R. A. (2008) Epigenomic consequences of immortalized plant cell suspension culture. *PLoS Biology,* 6**:** e302.

Tautz, D. (1989) Hypervariablity of simple sequences as a general source for polymorphic DNA markers. *Nucleic Acids Research,* 17**:** 6463-6471.

Teixeira, F., and Colot, V. (2010) Repeat elements and the *Arabidopsis* DNA methylation landscape. *Heredity,* 105**:** 14-23.

Tenaillon, M. I., Hollister, J. D. and Gaut, B. S. (2010) A triptych of the evolution of plant transposable elements. *Trends in Plant Science,* 15**:** 471-478.

Tester, M., and Langridge, P. (2010) Breeding technologies to increase crop production in a changing world. *Science,* 327**:** 818-822.

This, P., Jung, A., Boccacci, P., Borrego, J., Botta, R., Costantini, L., . . . Maul, E. (2004) Development of a standard set of microsatellite reference alleles for identification of grape cultivars. *Theoretical and Applied Genetics,* 109**:** 1448-1458.

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

Thompson, J. D., Higgins, D. G. and Gibson, T. J. (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Research,* 22**:** 4673-4680.

Thompson, M. M., and Olmo, H. P. (1963) Cytohistological studies of cytochimeric and tetraploid grapes. *American Journal of Botany,* 50**:** 901-906.

Tissier, A. F., Marillonnet, S., Klimyuk, V., Patel, K., Torres, M. A., Murphy, G. and Jones, J. D. G. (1999) Multiple independent defective suppressor-mutator transposon insertions in *Arabidopsis*: a tool for functional genomics. *The Plant Cell Online,* 11**:** 1841-1852.

Tittel-Elmer, M., Bucher, E., Broger, L., Mathieu, O., Paszkowski, J. and Vaillant, I. (2010) Stress-induced activation of heterochromatic transcription. *PLoS genetics,* 6**:** e1001175.

Torregrosa, L. (1998) A simple and efficient method to obtain stable embryogenic cultures from anthers *of Vitis vinifera* L. *Vitis,* 37**:** 91-92.

Tukey, J. W. (1949) Comparing individual means in the analysis of variance. *Biometrics***:** 99-114.

Turcich, M. P., Bokhari-Riza, A., Hamilton, D. A., He, C., Messier, W., Stewart, C.-B. and Mascarenhas, J. (1996) PREM-2, a copia-type retroelement in maize is expressed preferentially in early microspores. *Sexual Plant Reproduction,* 9**:** 65-74.

Untergasser, A., Cutcutache, I., Koressaar, T., Ye, J., Faircloth, B. C., Remm, M. and Rozen, S. G. (2012) Primer3 - new capabilities and interfaces. *Nucleic Acids Research,* 40**:** e115-e115.

Urbański, D. F., Małolepszy, A., Stougaard, J. and Andersen, S. U. (2012) Genome-wide LORE1 retrotransposon mutagenesis and high-throughput insertion detection in Lotus japonicus. *The Plant Journal,* 69**:** 731-741.

Van Houwelingen, A., Souer, E., Spelt, K., Kloos, D., Mol, J. and Koes, R. (1998) Analysis of flower pigmentation mutants generated by random transposon mutagenesis inPetunia hybrida. *The Plant Journal,* 13**:** 39-50.

Van Leeuwen, C., Friant, P., Choné, X., Tregoat, O., Koundouras, S. and Dubourdieu, D. (2004) Influence of climate, soil, and cultivar on terroir. *American Journal of Enology and Viticulture,* 55**:** 207-217.

Van Leeuwen, C., and Seguin, G. (2006) The concept of terroir in viticulture. *Journal of Wine Research,* 17**:** 1-10.

Van Steensel, B. (2011) Chromatin: constructing the big picture. *The EMBO journal,* 30**:** 1885-1895.

Vandesompele, J., De Preter, K., Pattyn, F., Poppe, B., Van Roy, N., De Paepe, A. and Speleman, F. (2002) Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. *Genome Biology,* 3**:** research0034.

Varshney, R. K., Graner, A. and Sorrells, M. E. (2005) Genomics-assisted breeding for crop improvement. *Trends in Plant Science,* 10**:** 621-630.

Velasco, R., Zharkikh, A., Troggio, M., Cartwright, D. A., Cestaro, A., Pruss, D., . . . Viola, R. (2007) A High Quality Draft Consensus Sequence of the Genome of a Heterozygous Grapevine Variety. *PLoS ONE,* 2**:** e1326.

Venturi, S., Dondini, L., Donini, P. and Sansavini, S. (2006) Retrotransposon characterisation and fingerprinting of apple clones by S-SAP markers. *Theoretical and Applied Genetics,* 112**:** 440-444.

Verdel, A., Jia, S., Gerber, S., Sugiyama, T., Gygi, S., Grewal, S. I. and Moazed, D. (2004) RNAi-mediated targeting of heterochromatin by the RITS complex. *Science,* 303**:** 672-676.

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**:** 366-376.

Vezinhet, F. ß., Hallet, J.-N. l., Valade, M. and Poulard, A. (1992) Ecological survey of wine yeast strains by molecular methods of identification. *American Journal of Enology and Viticulture,* 43**:** 83-86.

Vicient, C. M., Kalendar, R. and Schulman, A. H. (2001) Envelope-class retrovirus-like elements are widespread, transcribed and spliced, and insertionally polymorphic in plants. *Genome Research,* 11**:** 2041-2049.

Vidal, J. R., Gomez, C., Cutanda, M. C., Shrestha, B. R., Bouquet, A., Thomas, M. R. and Torregrosa, L. (2010) Use of gene transfer technology for functional studies in grapevine. *Australian Journal of Grape and Wine Research,* 16**:** 138-151.

Vielle-Calzada, J.-P., de la Vega, O. M., Hernández-Guzmán, G., Ibarra-Laclette, E., Alvarez-Mejía, C., Vega-Arreguín, J. C., . . . Herrera-Estrella, L. (2009) The Palomero genome suggests metal effects on domestication. *Science,* 326**:** 1078-1078.

Vigne, E., Komar, V. and Fuchs, M. (2004) Field safety assessment of recombination in transgenic grapevines expressing the coat protein gene of Grapevine fanleaf virus. *Transgenic Research,* 13**:** 165-179.

Visioli, G., Maestri, E., Polverini, E., Pavesi, A. and Marmiroli, N. (2013) AtL1 a non-LTR retrotrasposon fragment in the genome of *Arabidopsis thaliana* with homology to plants and animals. *American Journal of Plant Sciences,* 4**:** 806-816.

Vitte, C., Chaparro, C., Quesneville, H. and Panaud, O. (2007) Spip and Squiq, two novel rice non-autonomous LTR retro-element families related to RIRE3 and RIRE8. *Plant Science,* 172**:** 8-19.

Vivier, M. A., and Pretorius, I. S. (2002) Genetically tailored grapevines for the wine industry. *Trends in Biotechnology,* 20**:** 472-478.

Vos, P., Hogers, R., Bleeker, M., Reijans, M., Lee, T. v. d., Hornes, M., . . . Kuiper, M. (1995) AFLP: a new technique for DNA fingerprinting. *Nucleic Acids Research,* 23**:** 4407-4414.

Vukich, M., Giordani, T., Natali, L. and Cavallini, A. (2009) Copia and Gypsy retrotransposons activity in sunflower (*Helianthus annuus* L.). *BMC Plant Biology,* 9**:** 150.

Walbot, V. (1988) Reactivation of the Mutator transposable element system following gamma irradiation of seed. *Molecular and General Genetics MGG,* 212**:** 259-264.

Walbot, V. (1992a) Reactivation of *Mutator* transposable elements of maize by ultraviolet light. *Molecular and General Genetics MGG,* 234**:** 353-360.

Walbot, V. (1992b) Strategies for mutagenesis and gene cloning using transposon tagging and T-DNA insertional mutagenesis. *Annual Review of Plant Biology,* 43**:** 49-78.

Walbot, V. (1999) UV-B damage amplified by transposons in maize *Nature,* 397**:** 398.

Walbot, V. (2000) Saturation mutagenesis using maize transposons. *Current Opinion in Plant Biology,* 3**:** 103-107.

Walker, A. R., Lee, E., Bogs, B., McDavid, D. A. J., Thomas, M. R. and Robinson, S. P. (2007) White grapes arose through the mutation of two similar and adjacent regulatory genes. *The Plant Journal,* 49**:** 772-785.

Wang, H., Moore, M. J., Soltis, P. S., Bell, C. D., Brockington, S. F., Alexandre, R., . . . Soltis, D. E. (2009) Rosid radiation and the rapid rise of angiosperm-dominated forests. *Proceedings of the National Academy of Sciences,* 106**:** 3853-3858.

Waugh, R., McLean, K., Flavell, A., Pearce, S., Kumar, A., Thomas, B. and Powell, W. (1997) Genetic distribution of Bare1-like retrotransposable elements in the barley genome revealed by sequence-specific amplification polymorphisms (S-SAP). *Molecular and General Genetics MGG,* 253**:** 687-694.

Wegscheider, E., Benjak, A. and Forneck, A. (2009) Clonal variation in Pinot noir revealed by S-SAP involving universal retrotransposon-based sequences. *Am. J. Enol. Vitic.,* 60**:** 104-109.

Welsh, J., and McClelland, M. (1990) Fingerprinting genomes using PCR with arbitrary primers. *Nucleic Acids Research,* 18**:** 7213-7218.

Wessler, S. R. (1996) Plant retrotransposons: Turned on by stress. *Current Biology,* 6**:** 959-961.

Whatman FTA protocol BC01: Processing protocol for downstream AFLP analysis of sample DNA on an FTA® card, pp. GE Healthcare.

Wicker, T., Sabot, F., Hua-Van, A., Bennetzen, J. L., Capy, P., Chalhoub, B., . . . Schulman, A. H. (2007) A unified classification system for eukaryotic transposable elements. *Nature Reviews Genetics,* 8**:** 973-982.

Williams, J. G., Kubelik, A. R., Livak, K. J., Rafalski, J. A. and Tingey, S. V. (1990) DNA polymorphisms amplified by arbitrary primers are useful as genetic markers. *Nucleic Acids Research,* 18**:** 6531-6535.

Williams-Carrier, R., Stiffler, N., Belcher, S., Kroeger, T., Stern, D. B., Monde, R.-A., . . . Barkan, A. (2010) Use of Illumina sequencing to identify transposon insertions underlying mutant phenotypes in high-copy Mutator lines of maize. *The Plant Journal,* 63**:** 167-177.

Woodrow, P., Pontecorvo, G., Fantaccione, S., Fuggi, A., Kafantaris, I., Parisi, D. and Carillo, P. (2010) Polymorphism of a new Ty1-copia retrotransposon in durum wheat under salt and light stresses. *Theoretical and Applied Genetics,* 121**:** 311-322.

Xie, Z., Johansen, L. K., Gustafson, A. M., Kasschau, K. D., Lellis, A. D., Zilberman, D., . . . Carrington, J. C. (2004) Genetic and functional diversification of small RNA pathways in plants. *PLoS Biology,* 2**:** e104.

Xiong, Y., and Eickbush, T. H. (1990) Origin and evolution of retroelements based upon their reverse transcriptase sequences. *The EMBO journal,* 9**:** 3353.

Yamashita, H., and Tahara, M. (2006) A LINE-type retrotransposon active in meristem stem cells causes heritable transpositions in the sweet potato genome. *Plant molecular biology,* 61**:** 79-84.

Yano, S., Panbehi, B., Das, A. and Laten, H. (2005) Diaspora, a large family of Ty3-gypsy retrotransposons in *Glycine max*, is an envelope-less member of an endogenous plant retrovirus lineage. *BMC evolutionary biology,* 5**:** 30.

Yu, G.-X., and Wise, R. P. (2000) An anchored AFLP-and retrotransposon-based map of diploid *Avena*. *Genome,* 43**:** 736-749.

Yuan, Y.-W., and Wessler, S. R. (2011) The catalytic domain of all eukaryotic cut-and-paste transposase superfamilies. *Proceedings of the National Academy of Sciences*.

Zabkiewicz, J. (2008) A cleaner and greener New Zealand thanks to 2, 4, 5-T, science, and silicones. *Chemistry in New Zealand,* 72**:** 8.

Zerjal, T., Joets, J., Alix, K., Grandbastien, M.-A. and Tenaillon, M. (2009) Contrasting evolutionary patterns and target specificities among three Tourist-like MITE families in the maize genome. *Plant Molecular Biology,* 71**:** 99-114.

Zhang, Z., Schwartz, S., Wagner, L. and Miller, W. (2000) A greedy algorithm for aligning DNA sequences. *Journal of Computational biology,* 7**:** 203-214.

Zhu, J.-K. (2009) Active DNA demethylation mediated by DNA glycosylases. *Annual Review of Genetics,* 43**:** 143.

Zietkiewicz, E., Rafalski, A. and Labuda, D. (1994) Genome fingerprinting by simple sequence repeat (SSR)-anchored polymerase chain reaction amplification. *Genomics,* 20**:** 176-183.

Zohary, D. (2004) Unconscious selection and the evolution of domesticated plants. *Economic Botany,* 58**:** 5-10.

Zohary, D., and Spiegel-Roy, P. (1975) Beginnings of fruit growing in the Old World. *Science,* 187**:** 319-327.

Zottini, M., Barizza, E., Costa, A., Formentin, E., Ruberti, C., Carimi, F. and Lo Schiavo, F. (2008) Agroinfiltration of grapevine leaves for fast transient assays of gene expression and for long-term production of stable transformed cells. *Plant Cell Reports,* 27**:** 845-853.

Appendix A

Vector Diagrams

All vector diagrams produced using Geneious version 6 created by Biomatters (available from http://www.geneious.com).

A.1.1 pART27:GUS/GFP

This expression vector was used as a positive control for plant transformation (see 5.3.3).

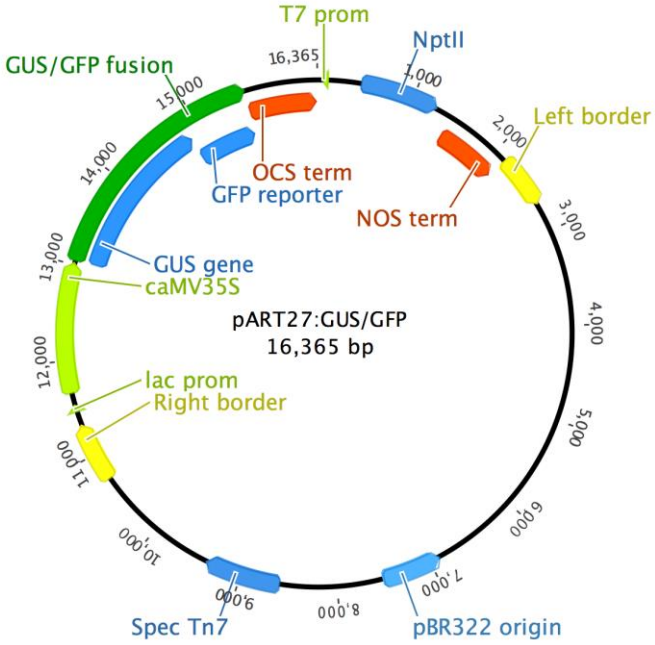


Figure A.1 Binary vector pART27:GUS/GFP

A.2 pART27: VvMYBA1

This expression vector contains the transcription factor VvMYBA1 preceded by a CaMV 35S promoter. Ectopic expression of VvMYBA1 in grapevine leaf tissue results in the production of the red-coloured anthocyanin pigment. This vector was used to quantify transformation efficiency in transient grapevine tissue transformations (see 5.3.2).

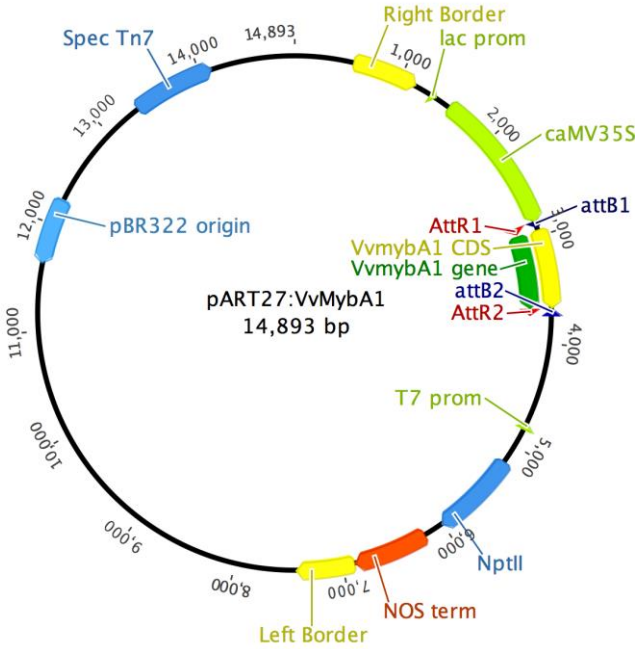


Figure A.2 Binary vector pART27: VvMYBA1

A.3 LTR-Reporter fusion vectors

These vectors contain the enhanced green fluorescent protein (EGFP) reporter gene preceded by LTR+UTR retrotransposon sequences. These vectors were used for transient and stable expression analysis of retrotransposon promoter regions (6.3.6).

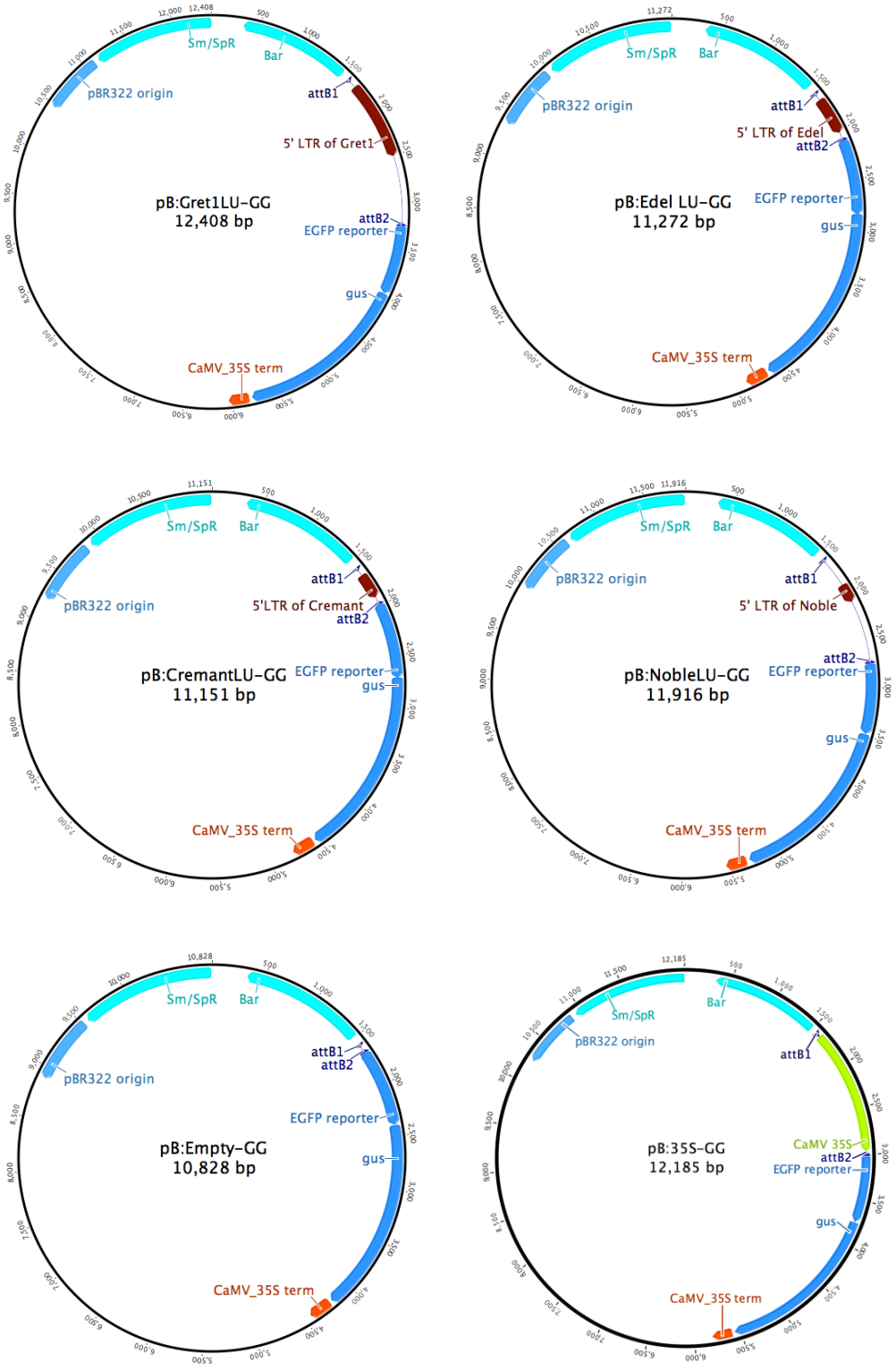


Figure A.3 Binary vectors used for in vitro promoter analysis of LTR regions of grape retrotransposons.

Appendix B

Supplementary Data

B.1 Alignments of grapevine LTR sequences

LTR insertions in the PN40024 grapevine genome were identified using MegaBLAST (ZHANG *et al.* 2000) and aligned with MUSCLE (EDGAR 2004). Bases differing from the consensus are highlighted in colour. Graphs above alignments display conservation (per cent) at each locus.

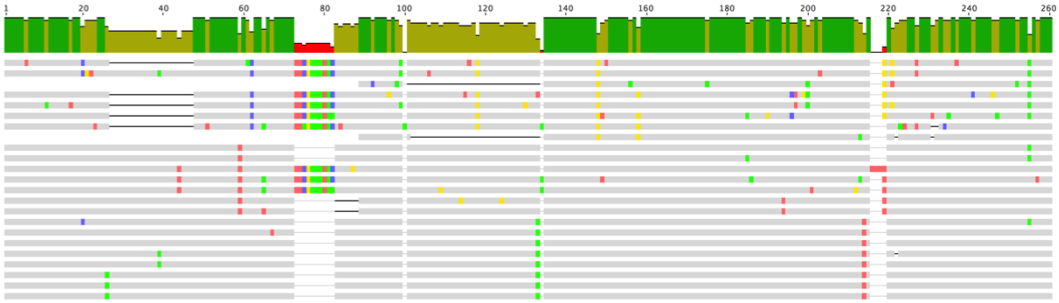


Figure B. 1 Alignment of *Cremant* LTRs in the grapevine (PN40024) genome.

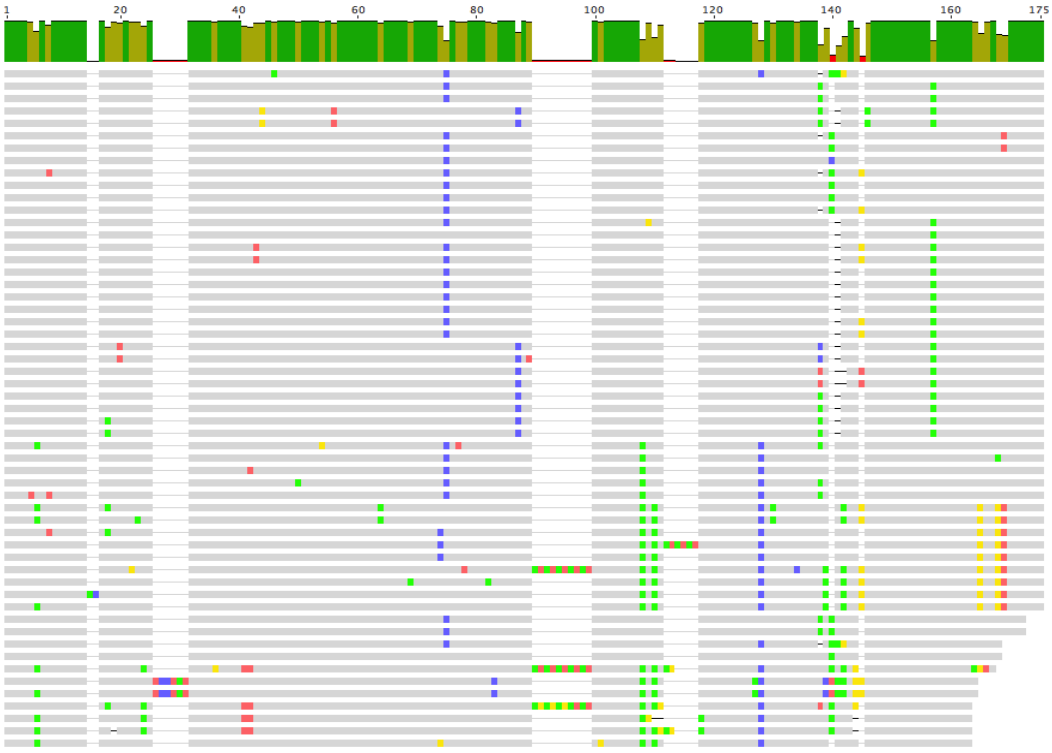


Figure B. 2 Alignment of *Tvv1* LTRs in the grapevine (PN40024) genome.

Figure B. 3 Alignment of *Gret1* LTRs in the grapevine (PN40024) genome.

Figure B.4 Alignment of *Edel* LTRs in the grapevine (PN40024) genome.

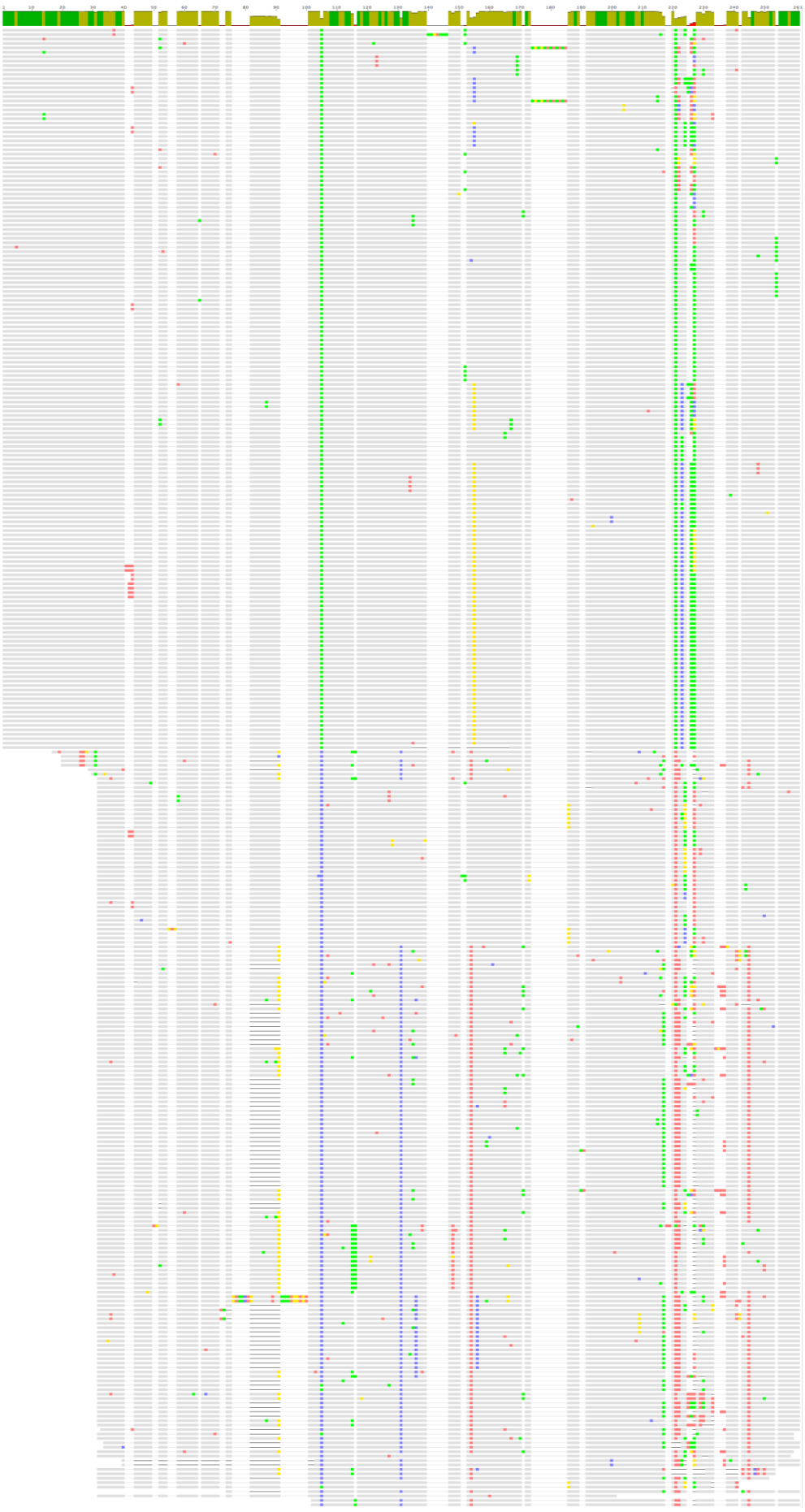
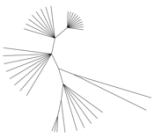
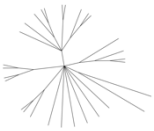
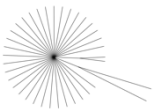
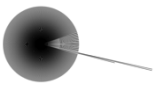
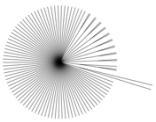
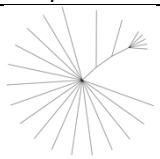
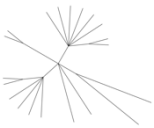
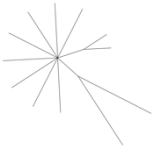
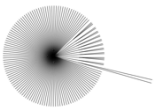
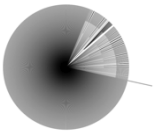
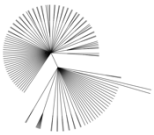
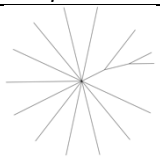
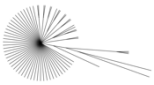
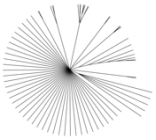
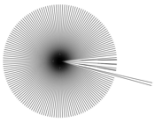
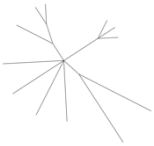
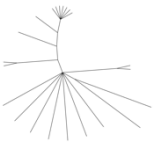
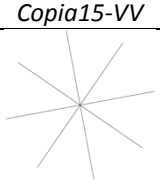
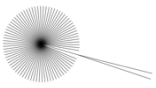
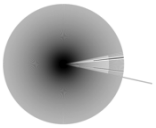
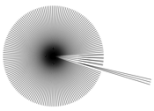
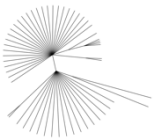
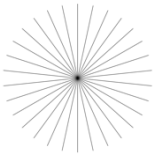
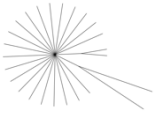
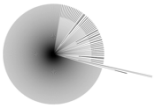
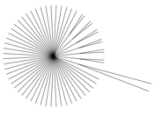
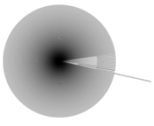
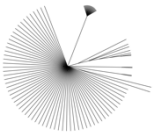
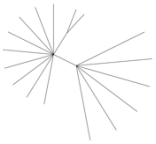
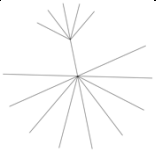
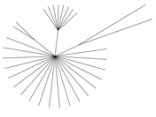
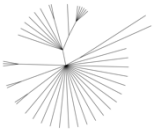
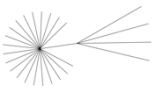
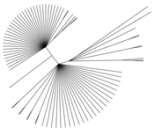
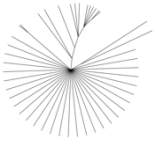


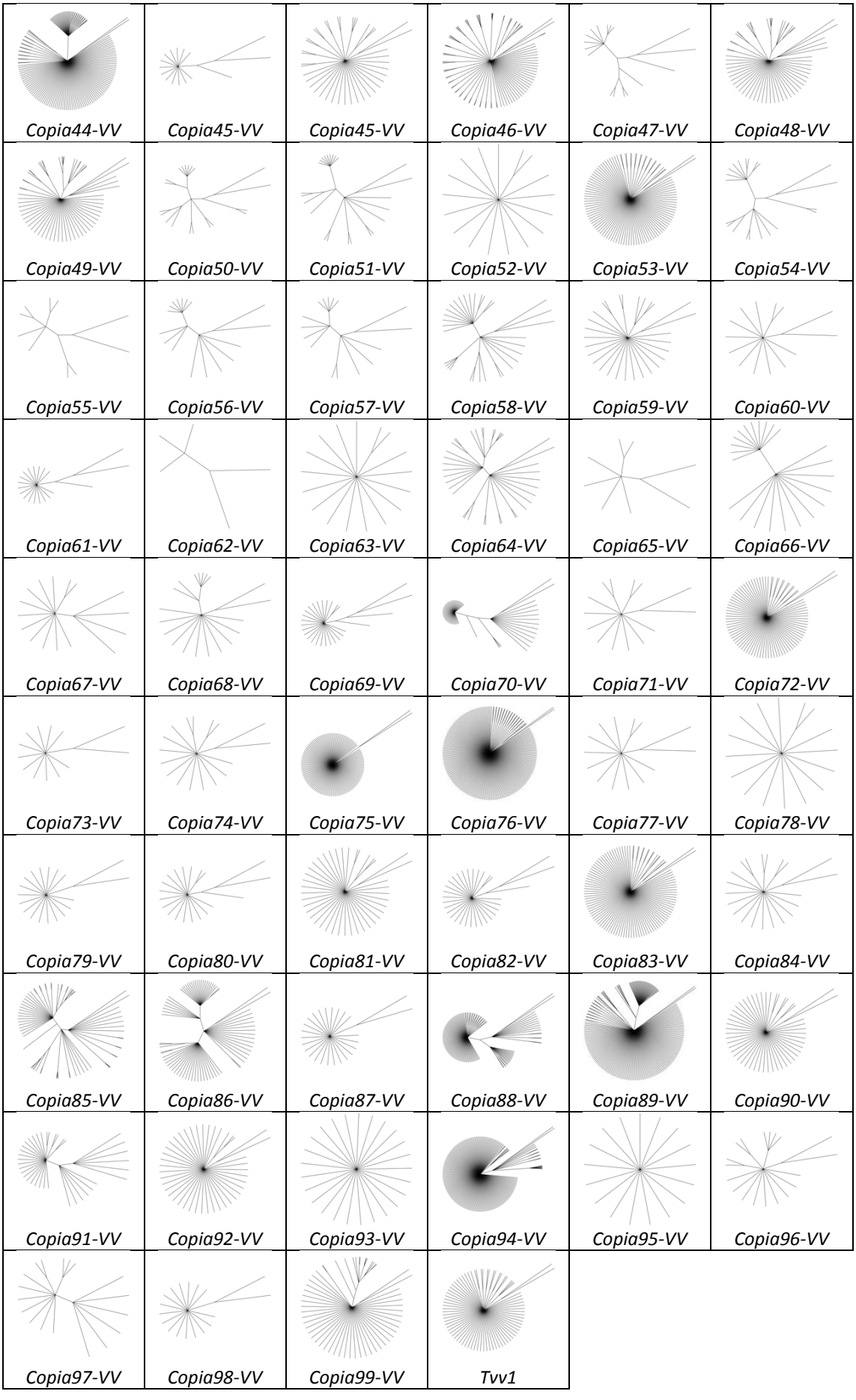
Figure B.5 Alignment of *Noble* LTRs in the grapevine (PN40024) genome.

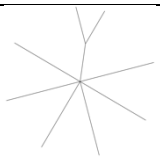
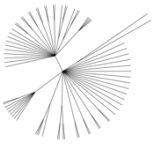
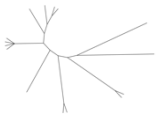
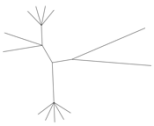
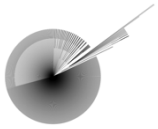
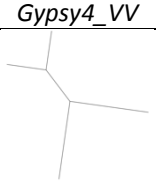
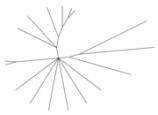
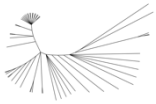
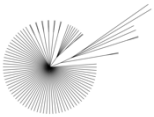
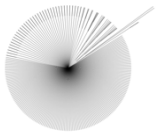
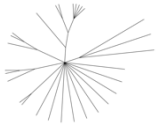
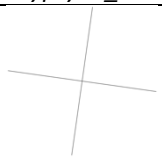
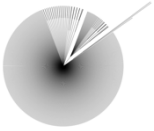
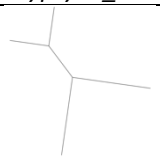
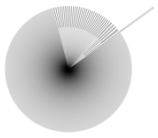
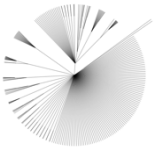
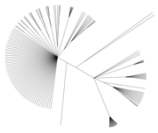
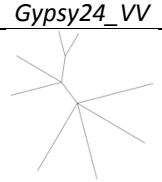
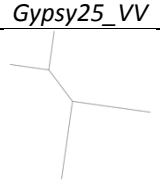
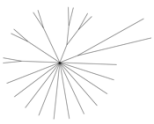
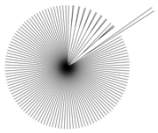
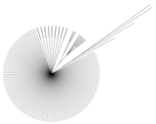
B.2 LTR phylogenetic trees

To visualise the historical proliferation of retrotransposon families within the grapevine genome phylogenetic alignments of LTR sequences were used (see 3.3.2). The cladograms presented here show the number and sequence similarity of repetitive LTR sequences for each defined retrotransposon family in the reference genome.

Table B.1 LTR sequence similarity cladograms for *V. vinifera* retrotransposon families. LTR sequences with >80% identity and >50% coverage within each of the retrotransposon families are included.

Ty1-*Copia* retrotransposons					
 *Copia1-VV*	 *Copia1A-VV*	 *Copia2-VV*	 *Copia3-VV*	 *Copia4-VV*	 *Copia5-VV*
 *Copia6-VV*	 *Copia7-VV*	 *Copia8-VV*	 *Copia9-VV*	 *Copia10-VV*	 *Copia11-VV*
 *Copia12-VV*	 *Copia13-VV*	 *Copia15-VV*	 *Copia16-VV*	 *Copia17-VV*	 *Copia18-VV*
 *Copia19-VV*	 *Copia20-VV*	 *Copia21-VV*	 *Copia22-VV*	 *Copia23-VV*	 *Copia24-VV*
 *Copia26-VV*	 *Copia27-VV*	 *Copia28-VV*	 *Copia29-VV*	 *Copia30-VV (Cremant)*	 *Copia31-VV (Edel)*
 *Copia32-VV*	 *Copia33-VV*	 *Copia34-VV*	 *Copia35-VV*	 *Copia36-VV*	 *Copia37-VV*
 *Copia38-VV*	 *Copia39-VV*	 *Copia40-VV*	 *Copia41-VV*	*Copia42-VV*	*Copia43-VV*


Copia44-VV
Copia45-VV
Copia45-VV
Copia46-VV
Copia47-VV
Copia48-VV
Copia49-VV
Copia50-VV
Copia51-VV
Copia52-VV
Copia53-VV
Copia54-VV
Copia55-VV
Copia56-VV
Copia57-VV
Copia58-VV
Copia59-VV
Copia60-VV
Copia61-VV
Copia62-VV
Copia63-VV
Copia64-VV
Copia65-VV
Copia66-VV
Copia67-VV
Copia68-VV
Copia69-VV
Copia70-VV
Copia71-VV
Copia72-VV
Copia73-VV
Copia74-VV
Copia75-VV
Copia76-VV
Copia77-VV
Copia78-VV
Copia79-VV
Copia80-VV
Copia81-VV
Copia82-VV
Copia83-VV
Copia84-VV
Copia85-VV
Copia86-VV
Copia87-VV
Copia88-VV
Copia89-VV
Copia90-VV
Copia91-VV
Copia92-VV
Copia93-VV
Copia94-VV
Copia95-VV
Copia96-VV
Copia97-VV
Copia98-VV
Copia99-VV
Tvv1

Ty3-*Gypsy* retrotransposons					
 *Gret1_VV*	 *Gypsy1_VV*	 *Gypsy2_VV*	 *Gypsy3_VV*	**2449*** *Gypsy4_VV*	 *Gypsy5_VV*
2391* *Gypsy6_VV*	**3296*** *Gypsy7_VV*	 *Gypsy8_VV*	 *Gypsy9_VV*	 *Gypsy10_VV*	**1284*** *Gypsy11_VV*
 *Gypsy12_VV*	**1015*** *Gypsy13_VV*	**2160*** *Gypsy14_VV*	 *Gypsy15_VV*	 *Gypsy16_VV*	**3857*** *Gypsy17_VV*
 *Gypsy18_VV*	 *Gypsy19_VV*	**3457*** *Gypsy20_VV*	 *Gypsy21_VV*	 *Gypsy22_VV*	 *Gypsy23_VV*
 *Gypsy24_VV*	 *Gypsy25_VV*	 *Gypsy26_VV*	 *Gypsy27_VV*	 *Gypsy28_VV*	 *Gypsy29_VV*
 *Gypsy30_VV*	 *Gypsy31_VV*	 *Gypsy32_VV*	 *Gypsy33_VV*	 *Gypsy34_VV*	

*****For those families with more than one thousand genomic copies only the copy number is given.

B.3 S-SAP comparison of multiple grape genotypes

Transposon insertion profiles were generated for 32 *Vitis* genotypes using the optimised S-SAP method (see 4.3.4). The profiles generated using primers specific to the Edel family of elements are given in the text (Figure 4.13). Those generated using *Gret1*, *Cremant*, and *Noble*-specific primers are shown here for comparison.

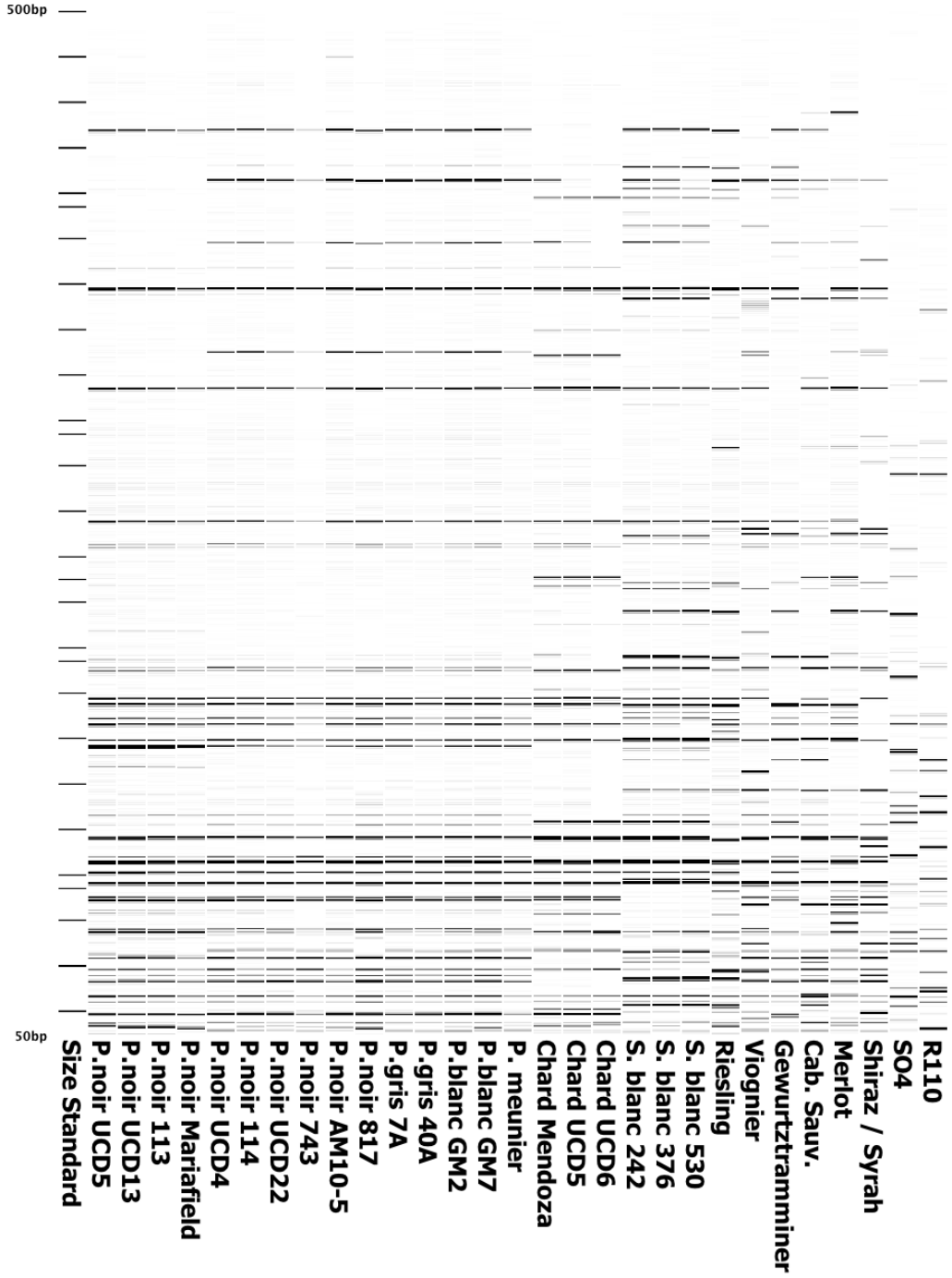


Figure B.6 S-SAP profiles generated by selective amplification of multiple genotypes with *Gret1* LTR primer and Mse(TG) primer.

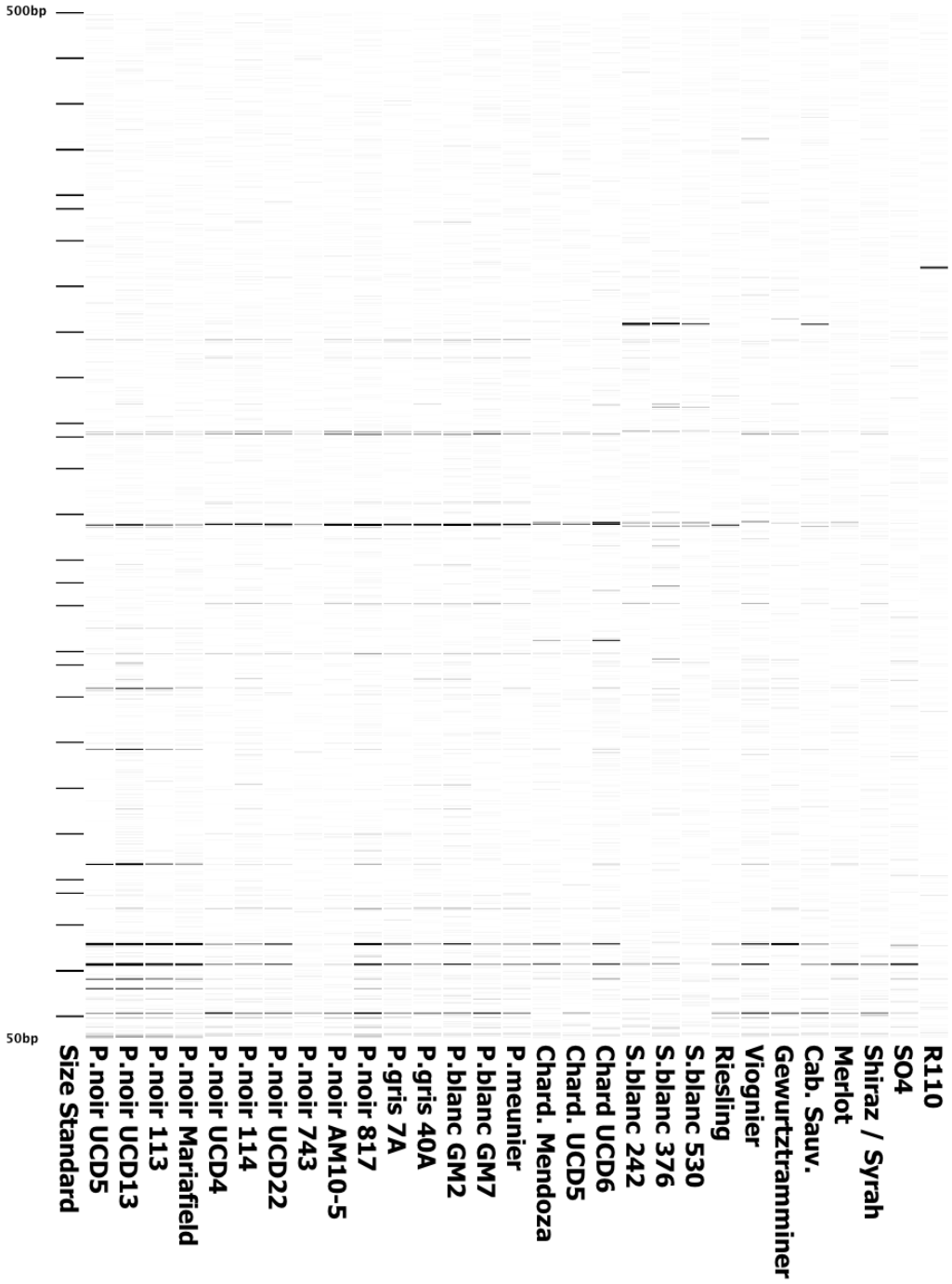


Figure B.7 S-SAP profiles generated by selective amplification of multiple genotypes with *Cremant* LTR primer and Mse(TG) primer.

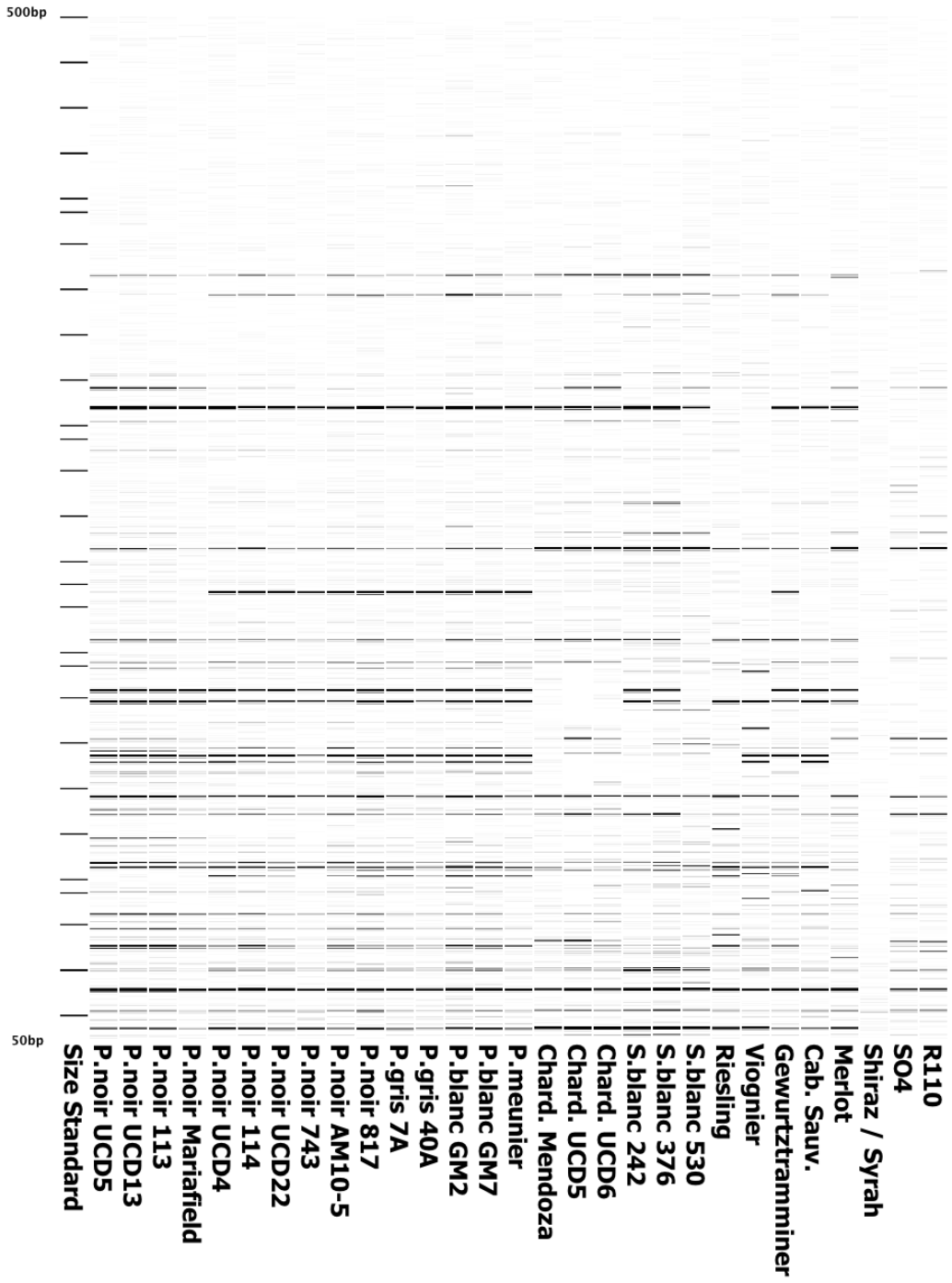


Figure B.8 S-SAP profiles generated by selective amplification of multiple genotypes with *Noble* LTR primer and Mse(TG) primer.

B.4 Tissue necrosis in transformed grapevine leaves

The inclusion of Pulse surfactant in *Agrobacterium* suspensions increased the efficiency of transformation of grapevine leaf tissue using a dip protocol (see 5.4.1). However, long dip periods and high Pulse concentrations resulted in tissue browning (Figure B.9). Any browning reduced the colour contrast between red anthocyanin-producing cells and the background untransformed tissue. In such cases, red cells could not be scored as positive transformations.

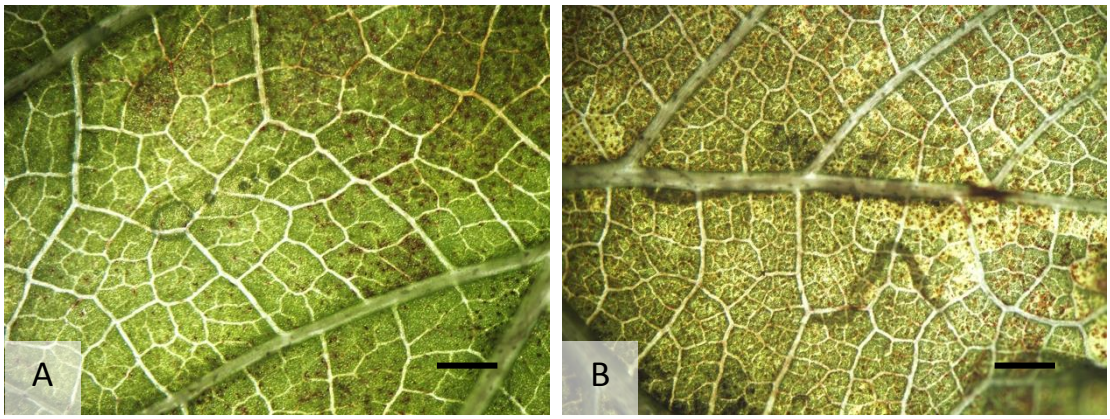


Figure B.9 Tissue necrosis in leaves treated with Pulse surfactant for longer periods or at higher concentrations. Dip transformation with *Agrobacterium* carrying pART27:VvMYBA1 (as described in 5.3.3) under differing conditions of Pulse concentration and time. **A.** Long duration: 20 minute dip duration, 0.1% (v/v) Pulse. **B.** High Pulse concentration: 10 minute dip duration, 1.0% (v/v) pulse. Scale bars represent 0.1mm.

B.5 Abnormal phenotypes of vines regenerated from somatic embryos

Images of regenerated vines appearing visually distinct to Pinot noir UCD5.

B.6 Time and fiscal cost of complete S-SAP analyses

In order to maximise the number of polymorphic retrotransposon insertions identified between any two genomic samples, S-SAP amplifications need to be performed with all adapter primers possessing 16 possible combinations of two selective bases. Calculations based on current reagent costs

Table B.2 (Following page) Cost calculations for comprehensive S-SAP amplification. The total time and financial cost associated with analysis of all insertions of four retrotransposon families in a single genome by S-SAP is given. Price estimates are based on quotes obtained in 2013 and are given in NZD.

Process	Reagent	Item Cost (NZD)	Samples processed per item	Cost per genotype	Contact time (hrs)	Non-Contact time (hrs)	Workflow
DNA Extraction	Nucleospin Plant II Kit (machery-Nagel)	$ 1,300.00	250	$ 5.20	5.00	1.00	Monday
	Qubit BR DNA quatification kit	$ 382.00	500	$ 0.76	0.50	0.00	
	EtBr-stained agarose gel	$ 500.00	12,500	$ 0.04	0.50	1.00	
Enzyme Digestion	MseI (NEB, 500U)	$ 244.00	500	$ 0.49	1.00	overnight	
	EtOH Precipitation	$ -	1	$ -	0.50	0.50	Tuesday
	EcoRI (Roche)	$ 50.00	1,000	$ 0.05	0.50	overnight	
	EtBr-stained agarose gel	$ 500.00	12,500	$ 0.04	0.50	1.00	Wednesday
Adapter ligation	MseI adapters (both strands)	$ 16.74	400	$ 0.04	0.50	0.50	
	EcoRI adapters (both strands)	$ 18.90	800	$ 0.02			
	T4 Ligase (Fermentas)	$ 307.00	1,000	$ 0.31	1.00	overnight	
	Denature ligase	$ -	N/A	$ -	0.25	0.50	
Pre-Amplification	Taq Polymerase	$ 200.00	250	$ 0.80	0.50	3.00	Thursday
	Mse(0) primer	$ 8.64	1,000	$ 0.01			
	EcoRI(0) primer	$ 8.64	1,000	$ 0.01			
	dNTP mix	$ 243.00	6,250	$ 0.04			
	EtBr-stained agarose gel	$ 500.00	12,500	$ 0.04	0.50	1.00	
SSAP (64 PCRs)	Taq Polymerase	$ 200.00	3	$ 66.67	2.00	3.00	
(=16 overhangs, 4 TEs)	Mse(XX) Primers (16)	$ 155.52	200	$ 0.78			
	Fluoro-primers (4)	$ 780.00	25	$ 31.20			
	dNTP mix	$ 243.00	39	$ 6.22			
	Capillary electrophoresis (16 samples)	$ 120.00	1	$ 120.00	1.00	overnight	Friday
	Pseudogel processing	$ -	N/A	$ -	2.00		
Total time (up to 24 genotypes processed simultaneously)					**16.25 hrs**	**1 week**	
Cost per genotype (NZD)				**$ 232.72**			

B.7 Whole-genome sequencing QC data

Data generated using the FASTQC software (ANDREWS 2010) from raw Illumina sequence reads. One library of five pooled regenerated plants (sequenced using one-quarter of one lane) and the data from the parental leaf material (sequenced using one full lane) are shown.

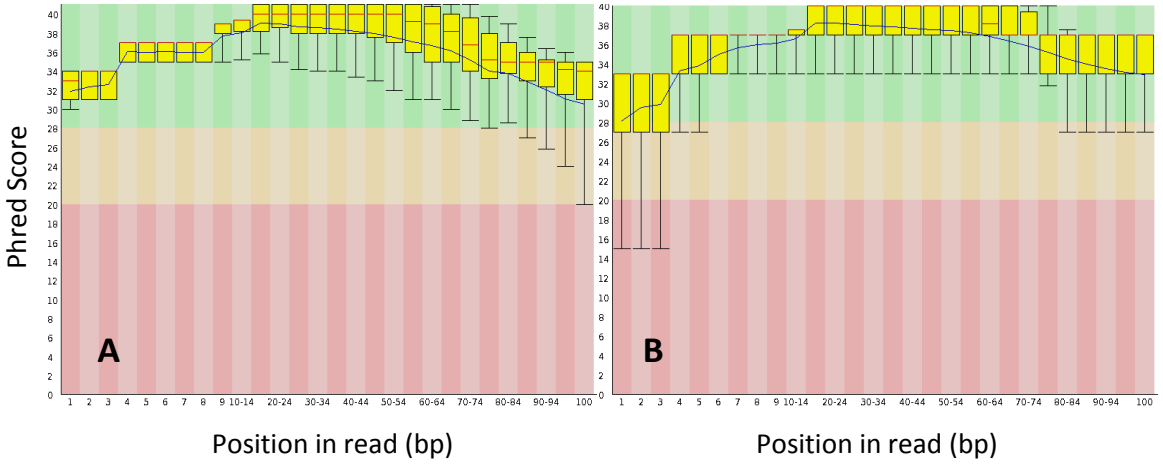


Figure B.10 Quality scores across all sequenced DNA fragments. A) Control 1 - sequences for left-hand reads from one of two lanes. **B)** P. noir UCD5 leaf (parental) - sequences for left-hand reads from one full lane.

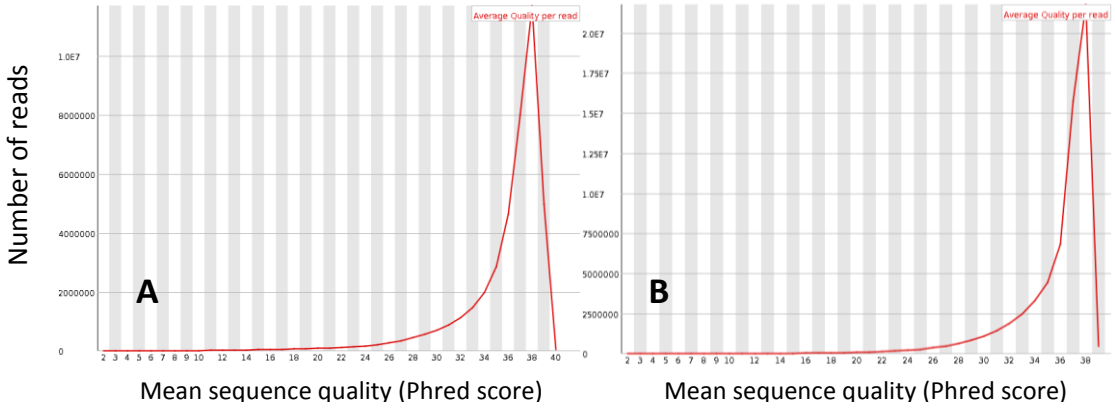


Figure B.11 Mean quality scores per sequence read. A) Control 1 - sequences for left-hand reads from one of two lanes. **B)** P. noir UCD5 leaf (parental) - sequences for left-hand reads from one full lane.

B.8 Characterisation of the insertion patterns of various grapevine TE families

For all TE families that had sufficient new insertion data to make the calculation possible (greater than 10 new insertions), insertions were found to be mainly in the Intergenic and intronic spaces of the genome (Figure B.12).

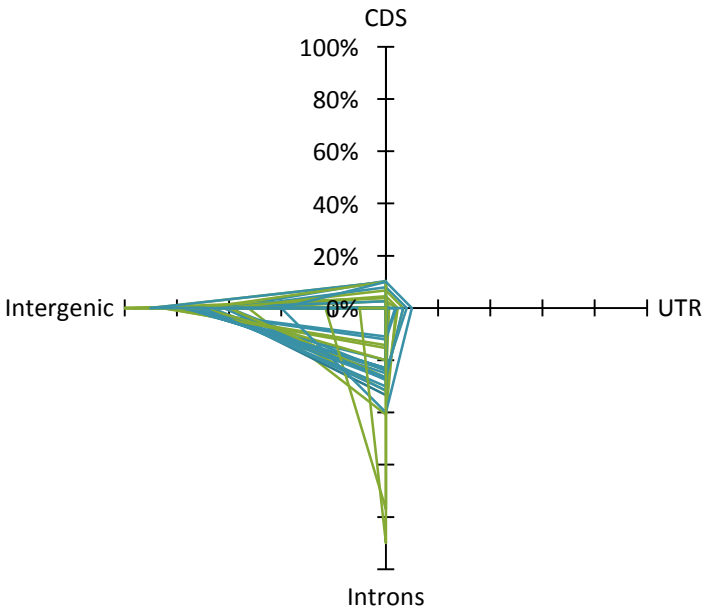


Figure B.12 Insertion site distributions of TE families. For all TE families that could be measured, the majority of novel insertions were located in the intergenic space or in introns. All TE families with greater than 10 novel insertions are plotted. Class I TEs are plotted in green and class II TEs are plotted in blue.

The four retrotransposon families selected for S-SAP profile analysis did not ultimately prove to be the most mutagenic in EC cultures. Whole-genome resequencing allowed activity of all predefined TE families to be characterised, without the selection bias of PCR-based approaches (Table B.3).

Table B.3 TE families displaying an average activity of at least one new insertion per individual.

Counts for total insertion numbers across all 20 of the vines regenerated from EC cultures are grouped by TE family. Those with at least 20 novel insertions (i.e. 1 new insertion per vine on average) are presented in descending order of transposition activity.

TE family	Novel insertions	CDS	UTR	Intron	Intergenic
Gypsy-23_VV	77	3	1	21	52
VLINE3_VV	76	2	3	31	40
MUDRAVI1	55	0	2	6	47
GYVIT1	44	2	2	14	26
VHARB-N1_VV	39	1	0	9	29
Gypsy-26_VV	38	0	0	9	29
Harbinger-3N1_VV	36	0	0	12	24
MuDR-5_VV	29	3	0	8	18
Gypsy17-VV	26	1	0	0	25
MuDR-18_VV	25	2	2	6	15
MuDR-21_VV	25	0	1	3	21
MUDRAVI2	25	0	0	5	20
VHARB4_VV	24	0	0	6	18
Gypsy-26_VV	21	0	0	3	18

Appendix C

Primer Catalogue

Table C.1 Record of oligonucleotide primers used in this study. Sequence modifications are entered in the format required by suppliers.

Primer Name	Sequence	T_m	Description	Section
Gret1 LTR rev(0) 6FAM	/56-FAM/CAGAGGGAGGGGTCCCTAC	63	Fluorescent dye-labelled primers for IRAP / REMAP / S-SAP from 5'LTR sequences.	4.3
Edel LTR Rev(0) VIC	VIC-CTAGGCTCCACAAAGCCCAACA	66		
Cremant LTR R(0) PET	PET-TGGACATTATCACTTAGTGGCTAAGCA	65		
Noble LTR R(0) NED	NED-AAGTCCTAACTTTGGCATTATTTCAACA	62		
EcoRI Adapter VIC	VIC-GACTGCGTACCAATTC	52		
Gret1LTR_DS-6FAM	FAM-CATCCATCTGGCAACTACGTGG	64	Fluorescent dye-labelled primers for IRAP / REMAP/ S-SAP from 3'LTR sequences.	4.3
EdelLTR_DS-VIC	VIC-GGTTGGGAATTCGGTTTAATTCCCT	64		
CremantLTR_DS-PET	PET-TTTCAGCCTTTATTCCACAACTT	59		
NobleLTR_DS NED	NED-ATCAATAAAGGAGTTCAGAAATTCTTCA	60		
Tvv1LTR_DS-6FAM	FAM-GAATTGAGAAATACCTTGATTCGGTTACA	63		
Gret1 R-Biotin	/5Biosg/CAGAGGGAGGGGTCCCTAC	63	Biotinylated S-SAP primers for amplification from 5'LTR sequences.	4.3
Edel R-Biotin	/5Biosg/CTAGGCTCCACAAAGCCCAACA	66		
Crem R-Biotin	/5Biosg/TGGACATTATCACTTAGTGGCTAAGCA	65		
Noble R-Biotin	/5Biosg/AAGTCCTAACTTTGGCATTATTTCAACA	62		
Gret1LTR_DS-bio	/5Biosg/CATCCATCTGGCAACTACGTGG	64	Biotinylated S-SAP primers for amplification from 3'LTR sequences.	4.3
EdelLTR_DS-bio	/5Biosg/GGTTGGGAATTCGGTTTAATTCCCT	64		
CremanLTR_DS-bio	/5Biosg/TTTCAGCCTTTATTCCACAACTT	59		
NobleLTR_DS-bio	/5Biosg/ATCAATAAAGGAGTTCAGAAATTCTTCA	60		
Tvv1LTR_DS-bio	/5Biosg/GAATTGAGAAATACCTTGATTCGGTTACA	63		
VvMYBA1 GWF	CACCATGGAGAGCTTAGGAGTTAGAAAG	60	Cloning of the VvMYBA1 gene	5.3.2
VvMYBA1 R	ATCAGATCAAGTGATTTACTTGTG	61		
Actin qPCRf	CTTGCATCCCTCAGCACCTT	63	Reference gene primers for grapevine qPCR (REID *et al.* 2006).	6.3.5
Actin qPCRr	TCCTGTGGACAATGGATGGA	61		
EF1α qPCRf	GAACTGGGTGCTTGATAGGC	61		
EF1α qPCRr	AACCAAAATATCCGGAGTAAAAGA	59		
GAPDH(m) qPCRf	TTCTCGTTGAGGGCTATTCCA	62		
GAPDH(m) qPCRr	CCACAGACTTCATCGGTGACA	63		
SAND qPCRf	CAACATCCTTTACCCATTGACAGA	61		
SAND qPCRr	GCATTTGATCCACTTGCAGATAAG	61		

qPCR Gret1 end F	GGCTTCCGCACCCCCTCTC	68	Primers for retrotransposon qPCR	6.3.5
qPCR Gret1 end R	TGATGATGGTCCGGCTGATAACAC	65		
qPCR Edel end F	TGCGCAGCTTCAATTGGTCTTCTA	66		
qPCR Edel end R	TTATTCGGGTCGGGTCGTCAAA	65		
qPCR Crem short2 F	ATCAAGCAAGTTGGGCATTC	60		
qPCR Crem short2 R	CACTTAACAGATGGGCATTATCA	59		
qPCR Noble short F	CATCCTCACCAAAGCTTTACC	60		
qPCR Noble short R	CCTAACTTTGGCATTATTTCAACA	59		
MseI linker 1	GACGATGAGTCCTGAG	53	Oligonucleotides for preparation of S-SAP adapters.	4.3.4
MseI linker 2	TACTCAGGACTCATC	47		
EcoRI linker 1	CTCGTAGACTGCGTACC	57		
EcoRI linker 2	AATTGGTACGCAGTCTAC	55		
HM linker1	GATCATGAGTCCTGCT	52		
HM linker2	CGAGCAGGACTCATGA	56		
MseI(0)	GATGAGTCCTGAGTAA	48	Preamplification primers for S-SAP	4.3.4
EcoRI(0)	GACTGCGTACCAATTC	52		
HpaII/MspI(0)	ATCATGAGTCCTGCTCGG	59		
Microsat-GA(C)	GAGAGAGAGAGAGAGAGAC	55	SSR primers for REMAP	4.3.3
Microsat-CT(G)	CTCTCTCTCTCTCTCTCTG	55		
MseI(XX)	GATGAGTCCTGAGTAAXX	48+	16 primers for selective S-SAP amplification. 'XX' is replaced with each permutation of 2 selective bases.	4.3.4
T7 F	TAATACGACTCACTATAGGG	53	Vector-specific primers for insert screening.	6.3.5
T7 R	TAGTTATTGCTCAGCGGTGG	61		
M13 F	GTAAAACGACGGCCAG	56		6.3.6
M13 R	CAGGAAACAGCTATGAC	52		
35S GWF	CACCGGCCGCTCGACGAATTAATTCCAATCCCA	63	Cloning of 35S promoter	6.3.6
35Sr	GCGTGTCCTCTCCAAATGAAATGA	56		
Gret1 GWF	CACCAGATTGTTCATCAAGGATACTAGTCAG	61	Amplification of 5'LTR + 5'UTR of canonical retrotransposons for Gateway cloning.	6.3.6
Gret1 UTRr	GTCATCCAGCCTTTTGCTGATGGAG	67		
Edel GWF	CACCTCGCATTTTAACATTACGGAGAGG	62		
Edel UTRr	TGCTCCCACTCAAACCTAACCCT	66		
Cremant GWF	CACCTTTATTAAATAATACTGATGCTACCAC	56		
Cremant UTRr	GATATATTTTCTTCAGGTATCTGG	55		
Noble GWF	CACCCCACATCATTGGTCAGTCCCTCCA	67		
Noble UTRr	GGACCGGAAAGATAGAAGAACACTTCTC	65		
GFP border check R	AACAGCTCCTCGCCCTTG	63	PCR confirmation of LTR-reporter construct	6.3.6

Appendix D

Recipes for Plant Tissue Culture Media

Table D.1 Ingredients for plant tissue culture media used in this study.

Medium	Reagent	Conc.	Amount	Stock conc.	Reference/ Notes
B Medium	**Initiation of grape embryogenic callus**				(PERRIN *et al.*
(1 litre)	MS salts & vitamins	1X	4.4g		2004)
	Sucrose		30g		
	Phytoagar		5g		
	BAP (100mM)	1mg/l	45µl	100mM	
	2,4-D (100mM)	1mg/l	45µl	100mM	pH5.8, autoclave
EG Medium	**Somatic embryo germination & *in vitro* plantlet growth**				
(1 litre)	MS salts & vitamins	0.5X	2.2g		
	Sucrose		30g		
	Phytoagar		7g		
	Activated charcoal		3g		pH5.8, autoclave
C_1^P Medium	**Embryogenic callus maintenance & proliferation (solid)**				(TORREGROSA
(1 litre)	MS basal salt mix	0.5X	2.2g		1998)
	MS micro elements	0.5X	0.5g		
	Casein hydrolysate		1g		
	Sucrose		30g		
	Gelrite		5g		
	T Vitamins		1ml	1000x	
	BAP (100mM)	1µM	10µl	100mM	
	2,4-D (100mM)	5µM	50µl	100mM	pH6.0, autoclave
T-Vitamins	**Added to C_1^P Medium**				(TORREGROSA
1000X stock	Myo-inositol		250mg		1998)
(5ml)	Nicotinic acid		5mg		
	Pyridoxine-HCl		5mg		
	Thiamine-HCl		5mg		
	Ca-Pantothenoate		5mg		
	D-Biotin		50µl	1mg/ml	
Long-term	**For potted vines**				
potting mix	Osmocote Exact Standard				
(1 litre)	12-14M (Everris)		3g		
	Hydraflo 2 (Everris)		1g		
	Dolomitic lime		3g		
	added to:				
	Composted bark		80% v/v		
	Pumice (grade 1-4mm)		20% v/v		
MS plates	**For general plant tissue maintenance & growth**				(MURASHIGE AND
(1 litre)	MS salts & vitamins	1x	2.4g		SKOOG 1962)
	Sucrose	3%	30g		
	Phytoagar		5g		pH5.8, autoclave
MS medium	**For *A. tumefaciens* resuspension**				(MURASHIGE AND
(*1 litre*)	MS salts & vitamins	1x	2.4g		SKOOG 1962)
	Sucrose	3% (w/v)	30g		pH5.8, autoclave
FM	**Embryogenic callus maintenance & proliferation (liquid)**				(BEN AMAR *et al.*
(1 litre)	NN salts & vitamins	1X	2.2g		2007)
	MES		0.5g		
	Maltose		20g		
	Glycerol		3.7ml		
	NOA	1mg/l	4.95ml	1mM	
	BAP	0.25mg/l	11.25µl		pH 5.8, autoclave
	Arabinogalactan (initially)		200µl	10mg/ml	

Appendix E

Supplementary Protocols

E.1 Annealing of adapter oligonucleotides

Double-stranded oligonucleotide adapters were prepared according to Syed and co-workers (SYED AND FLAVELL 2007). Briefly, the protocol is as follows:

Single-stranded oligonucleotide linkers were resuspended to a concentration of 100mM in TE buffer (linker sequences are given in Appendix C). Twenty microliters of each linker pair were mixed and incubated at 65°C for 10 min. The linkers were transferred immediately onto ice and 0.4µl of 1M magnesium acetate was added. After brief vortexing, samples were incubated at 37°C for 10 minutes and then at room temperature for a further 10 minutes. Adapters were stored at -20°C for up to one month.

E.2 Chemically competent *E. coli* transformation

Transformation of *E. coli* calls was performed according to (SAMBROOK AND RUSSELL 2001). Briefly, fifty microlitres of chemically competent E. coli DH5α cells were thawed on ice. To this, 1µl of ligation reaction was added. The cells were incubated on ice for 30 minutes, heat shocked at 42 °C for 45 seconds, and then returned to the ice. After 2 minutes 250µl of room temperature SOC medium was added to the cells. Cells were incubated at 37°C, 150RPM for 1hr and then 20µl and 200µl volumes were spread on separate LB plates with the plasmid-appropriate antibiotics. Plates were incubated for 16hrs at 37°C to allow colony growth.

E.3 Electro-competent *A. tumefaciens* transformation

Transformation of *A. tumefaciens* calls was performed according to (SAMBROOK AND RUSSELL 2001). Briefly, fifty microlitres of electro-competent *A. tumefaciens* GV3101 cells were thawed on ice and then transferred to an electroporation cuvette. To this, 1µl of purified plasmid was added. Electroporation of cells was performed with a MicroPulser (Bio-Rad) using the *Agrobacterium* selection, and cells were returned directly to ice. After two minutes, 500µl of room temperature SOC medium was added to the cells. Cells were incubated at 28°C, 150RPM for 1hr and then 20µl and 200µl volumes were spread on separate LB plates with the plasmid-appropriate antibiotics. Plates were incubated for 48hrs at 28°C to allow colony growth.

E.4 Colony PCR

Colony PCRs were used to screen bacterial colonies following transformation experiments for the presence of the vector containing the desired insert in the correct orientation. Generally, one vector-specific primer and one insert-specific primer are used, with the pair selected to amplify only insertions in the desired orientation.

Single colonies are picked by touching them with a sterile pipette tip. The pipette tip is then swirled in prepared PCR mix in a 0.2ml PCR tube and discarded. The reaction conditions (Table E.1) involve an extended initial incubation at 95°C in order to completely lyse all bacterial cells.

Table E.1 PCR reaction conditions for a colony PCR

PCR Mix	**Vol. (µl)**	
dH_2O	15.5	
10X CL PCR Buffer	2	
dNTPs (2.5 mM each)	1.6	
Primer1 (10µM)	0.4	
Primer2 (10µM)	0.4	
Taq Polymerase (Qiagen)	0.1	
Template	Colony pick	
Total Vol.	20	
PCR Conditions:		
Temp (°C)	Time	Cycles
95	10 min	
95	30 sec	35X
50	30 sec	
72	1 min	
72	5 min	

Ten microliters of each amplification product were run on a 1% (w/v) agarose TBE gel containing ethidium bromide (0.5µg/ml) at 100V for 20-30 minutes.

E.5 Sanger sequencing

Samples were prepared for chain-termination sequencing by combining 15-20ng of specific PCR product with 5pmol of the sequencing primer. Sequencing reactions were performed using the BigDye Terminator v3.1 kit (Applied Biosystems) according to the manufacturer's recommendations. Amplification products were analysed by capillary electrophoresis on an ABI 3130xl automated sequencer (Applied Biosystems). Relevant parameters are given in Table E.2.

Table E.2 Run parameters for ABI 3130xl capillary sequencer

Parameter	Value
Polymer	POP7
Array length	50cm
Run temperature	60°C
Pre-run voltage	15kV
Pre-run time	180 sec
Injection voltage	1.6kV
Injection time	15 sec
Data delay time	480 sec
Run voltage	8.5kV
Run time	6000 sec

E.6 GUS reporter assays

Two days after infiltration with *A. tumefaciens*, plant tissues were placed in 50ml falcon tubes. Enough GUS stain was added to completely cover the tissue. Tissue was then vacuum infiltrated for 2 x 2 minutes so that tissue appeared translucent. The tubes were then sealed and samples were incubated overnight at 37°C. Chlorophyll was removed by soaking in methanol for 2 hours and then in ethanol overnight at room temperature with gentle agitation.

Table E.3 GUS stain composition

Reagent	Concentration
Phosphate buffer (pH 7.0)	50mM
Triton X-100	0.01% (v/v)
X-gluc*	0.03% (w/v)
$K_4Fe(CN)_6$	1mM
$K_3Fe(CN)_6$)	1mM

*Dissolve X-gluc in approximately 50µl DMSO before adding

Wordle (www.wordle.net) generated using the entire text of this document. Font size corresponds to word frequency.