Isolation, characterisation and expression of *Ty1-copia* retrotransposons in *Agave tequilana*

Ishtiaq Khaliq

Doctor of Philosophy

University of Sussex
April 2009

I hereby declare that this thesis has not been and will not be submitted in whole or in part to any other university for the award of any other degree.

Signature_____

Ishtiaq Khaliq **University of Sussex**

Isolation characterisation and expression of *Ty1-copia* retrotransposons in *Agave tequilana*

Abstract

Agave tequilana is native to Mexico and famous for its use in the production of alcoholic beverage tequila. Due to the growing demands of tequila it has been intensively cultivated through asexual means resulting in a narrow genetic variability. The crop is extremely prone to insects and diseases and a breeding programme for selection and conservation of the crop is required. However poorly understood classification of agave and limited availability of molecular data is a big hurdle in establishing a useful breeding programme. Retrotransposons are mobile genetic elements that can replicate through a copy and paste mechanism occupying large proportions of genomes in short periods of time. They can play a vital role in the organisation and evolution of plant genomes and retrotransposon based molecular markers can be used as powerful molecular tools. Retrotransposons can play a key role in understanding *Agave tequilana* genetics as well as its relationship with other agave species and cultivars. In the work described here *Ty1-copia* retrotransposon have been isolated and characterised, retrotransposon based molecular markers have been used to evaluate the asexual genetic diversity in different vegitatively propagated mother and daughter plants. The phylogenetic relationship of these sequences revealed that *Ty1-copia* retrotransposons are heterogeneous in nature and they might have been actively replicating in recent past. The copy number of *Ty1- copia* retreotransposon was also evaluated and it was found that a large proportion of agave genome is occupied by these elements. *Ty1-copia* retrotransposons were also found to be polymorphic in agave tissue culture lines suggesting that they might be activated under the effect of stressed conditions. The findings of this study will help in understanding the genetics of *Agave tequilana*, and will provide a basis for further research on retrotransposon mediated asexual genetic variability in agave and other clonally propagated plants in general. It will also help us understand the activity of retrotransposons in the genome of agave in unusual environmental conditions.

Table of contents

CHAPTER 2

General Materials and Methods

CHAPTER 5

Retrotransposon expression and somaclonal variation in *Agave tequilana*

CHAPTER 6

Concluding Remarks

Acknowledgements

First of all I would like to take this opportunity to thank my supervisor Dr .Steve Pearce for his supervision and guidance. He has been very helpful to me throughout my time at Sussex. I will never forget his support during the writing of this thesis. I would also like to thank Professor Derek Lamport and others in the department for their help and support during the course of my stay at Sussex.

My parents always worked very hard to bring us up as responsible members of society and without their affection, encouragement and support I could not have achieved many goals of my life including this project. Thanks are not enough to express my feelings for them.

I would like to say thanks to my brothers and sisters who have always been there for me. I would like to mention my son Hanzalah's name who has been a refreshing change in my life.

Many thanks to all of my friends who have never let me down .Their moral and financial support got me trough this big task.

Last but not least thanks to my wife Saeeda for her continues help and support during the course of my PhD. Without her efforts I could not have completed this thesis.

CHAPTER 1

General Introduction

1.1 Agave and its Biology

1.1.1 History, conventional uses and distribution of agave

The agave was recognized as the "century plant" by Anglo-American while it has been known as "mescal" by Mexicans (Valenzuela-Zapata and Nabhan, 2003). Agave has been a part of the human life since the colonization of North American continent. The archeological evidence suggests human interaction with agave as early as 9000 years ago. Many agaves have been domesticated and cultivated prehistorically by indigenous communities and Native American tribes as a source of food, fiber and tools and genetic characteristics of the plant have been slowly influenced by human activity (Sheldon 1980; Gentry 1982; Colunga and Maypat 1993). Apart from the use of different parts of the plant as food the long fibers from the leaves of some species have also been used to make ropes and baskets. Different agaves have been widely used as medicines due to their antiseptic, anti inflammatory, diuretic and laxative properties (Verastegui et al., 1996). The fermentation of agave sugars has been used to produce alcoholic beverages and one of these traditional alcoholic beverages called *pulque* has been famous for its use in religious rituals (Miller and Taube, 1993). Some rural communities in Mexico still utilize agaves for the production of fiber, medicines and construction materials.

Although the geographic center of origin of agave family is Mexico where large populations of agave are found, but they are widely distributed from southern parts of Canada to the Andean highlands of northern South America (Gentry, 1982; Valenzuela-Zapata and Nabhan, 2003). The cultivation and use of different species of agave spread in Europe after the arrival of the Spanish conquistadors in 16th century. Agave is also found in other parts of the world because it has been transported to different parts of the world by humans and it can adapt to a wide variety of environmental conditions (Badano and Pugnaire, 2004). Agave plants are very successful in semi–arid regions and comprise a significant proportion of the flora of these regions but they can

also be found in a wide range of habitats and diverse environments for example deserts, grasslands, oak-pine woodlands and rocky steep surfaces (Nobel, 1988).

1.1.2 Taxonomy of agave

Agave has traditionally been classified with plants containing inferior ovaries and is considered to be a part of the Amaryllidaceae family while plants with superior ovaries for example Yucca are put together into a separate family Liliaceae (Bentham and Hooker, 1883; Engler and Prantl, 1888). In early 1930s the traditional classification of agave was revised and the plants with fibrous leaves like Yucca and agave were combined into a new family, the Agavaceae (Hutchinson, 1934). Two types of taxonomic systems are in use for the classification of agave, Cronquist's system which classifies the Agavaceae family as an expanded family of 18 genera (Cronquist, 1981), while the other system defined the Agavaceae family as a narrow family consisting of only two tribes, the Yucceae and Agavaceae and 9 genera of plants (Dahlgren and Yeo, 1985).

Traditionally the phylogenetic analysis based on the morphological and anatomical characters have been used, however the development of new technologies have facilitated a more reliable classification of Agavaceae family, for example chloroplast DNA restriction site analysis (Bogler and Simpson, 1995) and internal transcribed spacer region of the ribosomal DNA supported the narrow definition of the Agavaceae family. The Dahlgren classification was also thought to be suitable by the sequence analysis of the large subunit of ribulose-1,5- biphosphate carboxilase (Duvall et al., 1993). There are different opinions about the estimated age of the Agavaceae family, and it is thought to be between 20-26 million years (Good-Avila et al., 2006) and 35 million years (Wikstrom et al., 2001). The genus *Agave* has been grouped with three other genera and collectively called *Agave sensulato*. However the *Agave* genus itself is termed as *Agave sensu stricto* (Eguiarte et al., 2000).

The number of species in Agavaceae family recognized so far is approximately ≈293 species out of which ≈ 208 species belong to the *Agave sensulato* and if 49 Yucca species are also included the number of species becomes 257. The *Agave sensulato* is considered to be the youngest in the Agavaceae family (Good-Avila et al., 2006) but it contains 208 species and 166 species out of these 208 belong to the agave genus *Agave sensu stricto*. The high number of species in the *Agave sensulato* suggests a great diversification of this genus in a short evolutionary time scale.

A rapid speciation has specifically occurred in the *Agave sensu sticto* (*Agave* genus) with two elevated peaks of speciation, one between 6 and 8 Mya and the other one between 3 and 2.5 Mya. The first speciation peak coincides with the dry environmental conditions in central Mexico 8-6 Mya which might have triggered the diversification. The greater adaptation of Agavaceae family to dry and arid habitats might also have a link to the first speciation peak. The *Agave sensu stricto* comprises of two subgenera namely Littaea and Agave, and the subgenus Agave contains 113 species (Good-Avila et al., 2006). However the precise number of Agave species and their varieties is unknown because of the high number of species and varieties as well as the complex polyploidy variations in some of them (Palomino et al., 2003). Additionally the limited reliability on the old classification methods and limited availability of data for the modern molecular biological techniques (Garcia-Mendoza and Chiang, 2003) as well unavailability of the floral structures due to long life span of many species like *Agave tequilana* (Gentry, 1982; Valenzuela-Zapata, 1997) cause problems in the classification of different varieties of the same species.

1.1.3 Morphology of agave

Agaves are xerophytes and highly successful in harsh environmental conditions. Their biological success in such difficult and dry conditions is due to their ability to retain high levels of water in their leaves. Agaves grow in the form of a rosette structure with leaves growing from a short central shoot in an organized manner. The new leaves grow from the upper part of the shoot and the basal parts of the leaves form the central head of the plant which is called the “pinna”. The pinna is used as the "store house" of the plant for the products of photosynthesis in the form of sugars (Gentry, 1982) which are subsequently used industrially for the production of tequila. The weight of the pinna can vary from 25 kg to 75kg but some times pinnas as heavy as 200 kg have been reported. The nutrients stored in the central head (pinna) of the plant are supplied to the other parts like roots, rhizomes and floral stalk. One of the main morphological characteristics of agave is the long succulent fleshy leaves of varying sizes ranging from 20cm to 2m (Nobel, 1988; Valenzuela-Zapata and Nabhan, 2003). The succulent leaves have high water storage capacity to cope with the prolonged drought periods (Smith and Nobel, 1986). The agave plant has a cuticle which covers the leaves and is impermeable which reduces transpiration (Smith and Nobel, 1986). These two important morphological factors increase water storage capacity which

is very important for the adaptation and success of agave in dry and arid habitats. The leaves vary in colour from bluish to green and grey while some species have yellow stripes on the leaves for example some varieties of *Agave americana* (Gentry, 1982). Agaves usually have spiky leaves but in some varieties the spikes are absent. These spikes can serve as a defense against herbivores. Agaves also develop an inflorescence which usually very tall except for some ornamental varieties. The central inflorescence can range between 2-10 meters but most of the commercial agave varieties of *Agave tequilana* are harvested before the formation of inflorescence.

1.1.4 Proliferation of agave in the dry and arid habitats

Although agaves are adaptable to a variety of environments, the scarcity of water seems to be the key factor involved in the evolution of agave genus and its adaptation to dry arid habitats (Gentry, 1982). Apart from the morphological characteristics such as the succulent leaves and presence of a thick cuticle (Nobel, 1991; Nobel et al., 1992) a carbon metabolism-related drought-adaptation also plays a major part in their adaptation to semi-arid and arid habitats. Members of Agavaceae family are among the plant species which utilize the Crassulacean Acid Metabolism (CAM) biochemical pathway that permits the fixation of atmospheric CO_2 for photosynthesis during the night when temperatures are significantly lower than day time temprature (Winter and Smith, 1996).

CAM is common in the plants of dry and arid environments like deserts, during the day time when temperatures are high the CAM plants close their stomata, thus reducing the loss of water through transpiration while at nights the stomata are opened for the uptake of CO_2 which is fixed to a phosphoenolpyruvate (PEP) with the help of PEP-carboxylase (PEPC) by producing malate which is then stored in the cell vacuole with the help of an H^+-ATPase pump (Dodd et al., 2002). At the start of the day the leaf cell prepares for the diurnal phase of CAM and photosynthesis: the stomata close, PEPC activity is diminished, and malate passively exits from the vacuole to the cytosol where it is enzymatically de-carboxylated. The available CO_2 is fixed via the Calvin cycle that leads to carbohydrate synthesis.

Although most of the plant species that are adapted to drought use CAM but many species can behave differently under specific environmental and physiological factors. So the common diurnally CO_2 uptaking species can switch to CAM when there is a drought or in a condition of

heat stress, similarly CAM species can also make changes to their regular norms when the water is abundant by reducing total uptake of CO_2 (Dodd et al., 2002). Species such as agaves have additional special morphological processes for example the changes in the opening of stomata and adaptations for example succulent leaves and expanded vacuoles that can accommodate the increased accumulation of organic acids (Nobel, 1988). Another reason for the adaptation of agave in semi arid and arid regions is the accumulation of uncommon carbohydrates called fructans (Hendry, 1993). The fructans can help desert plants species during the environmental stresses like cold, heat, and drought, where they play a vital role in stabilization of cell membrane (Ritsema and Smeekens, 2003; Saldana, 2006). So the succulent leaves with impermeable cuticle, CAM and accumulation of fructans seem to be the driving forces for the evolution of in harsh environmental conditions and their adaptation to the dry, arid and semi-arid habitats.

1.1.5 Life cycle of agave

Most of the species in Agavaceae family have a long life span; however the life expectancy varies with the species and ranges from 5-6 years to 20 years (Gentry, 1982). Many environmental conditions and soil fertility can affect the life expectancy of different agave species (Nobel, 1988) for example in *A. tequilana* the warmer conditions can facilitate high metabolic activities and shorter life-span. Agaves can reproduce through sexual and asexual means. However in species of commercial importance such as *A. tequilana,* asexual propagation is commonly in practice due to long life span of most of agave species. During asexual reproduction vegetative offshoots are produced via underground rhizomes of the mother plant (Tissue and Nobel, 1988). These offshoots develop into daughter plants that will remain attached to the mother plant and nutrients will be translocated to them until they start their own photosynthesis and eventually start their independent life. Asexual propagation is common in agaves and has helped the local proliferation of some species in new areas; however it is the main cause of low genetic variability in agave species and their vulnerability to diseases and pathogens in the cultivated crop (Valenzuela-Zapata and Nabhan, 2003). Asexual propagation is also possible through bulbils which develop from the cells at the flower stem of some agave species (Arizaga and Ezcurra, 1995) but this is uncommon.

Agaves reproduce sexually by producing an inflorescence which can be 2-10 meters tall. This inflorescence is produced from the center of the central head of the plant by increased metabolic activities consuming all of the accumulated sugars. However after the development of the

inflorescence, pollination and seed production occur in a short period of time after which the plants usually die (Gentry, 1982; Nobel, 1988). Large number of seeds can be produced with high germination rates but the survival rate of the germinated seedlings is very low with most of them dying with 8-9 days of germination (Arizaga and Ezcurra, 2002).

Many pollinators pollinate the agave and bats are suggested to be the most important as co-evolution of many agave and bat species have been reported (Arita, 1991; Slauson, 2001). Although sexual reproduction is very important for the generation of the genetic diversity of agave but due to its long time period, and consumption of the commercially important sugars, it is not usually allowed to occur in the cultivated agave species.

.

1.1.6 *Agave tequilana* Weber var. azul (the blue agave) and Tequila.

Agave tequilana is famous for its use in the production of alcoholic beverage tequila which is the most profitable product of agave cultivars. Historically tequila was one of the alcoholic drinks produced from the different agave species and known as "*vino-mescal de tequila"*. Tequila has been imported to the United States of America since the end of 19th century but a sudden rise in its popularity in late 1970s doubled the consumption of tequila in the US and eventually it has became globally famous with its sale in more than 40 countries. Due to the growing demand of tequila the cultivation of other agave species for the production of food, fiber and drinks has been ignored and replaced by plantations of blue agave for the production of tequila (Valenzuela-Zapata and Nabhan, 2003). At the present time tequila is one the most important exports of Mexico consisting of 400 brands made by 30 different companies. Between 2002 and 2007 a 46% increase in the import of tequila has been observed. Until 1970s four different species were mixed to get tequila, however in last 20 years the production of tequila has been restricted to only one variety of *Agave tequilana* species, the Weber azul. On the other hand the name tequila is protected according to the international agreements and the variety used for its production can only be cultivated in certain states of Mexico including Jalisco which is the major producing state (SCFI, 1997), now blue agave constitutes about 99% population of agave in Jalisco (Dalton, 2005; Valenzuela-Zapata and Nabhan, 2003). Due to the growing demand of tequila and long life span of agave , blue agave has been vegetatively propagated resulting in the formation of large population of clonal plants spreading over 120,000 acres in a few years time. The clonal propagation reduced the genetic variability and largely homogeneous crop was devastated by bacterium (*Erwinia carotovora*) and a fungus (*Fusarium oxysporum*) in mid 90s. These

outbreaks increased the prices of tequila but also increased an interest in the genetic research of *Agave tequilana* (Dalton, 2005; Hidalgo et al., 2004).

1.1.7 The blue agave genetics research

The loss of blue agave to disease has led to a growing concern about the genetic diversity of the crop and resulted in an increased interest in genetic research of the plant. In recent years the investigations of genetic diversity within blue agave and related tequilana varieties and wild relatives have started aiming to identify varieties of blue agave and establish breeding programmes that can address the genetic variation and improve the resistance to pathogens (Dalton, 2005). It was initially suggested that the level of genetic diversity among the natural population of blue agave was very low as the randomly amplified fragment polymorphism analysis revealed a very low level of polymorphic bands (Vega et al., 2001). In contrast, great genetic variability was found in blue agave by the same group five years later using AFLP (Vega et al., 2006). The contradiction in the results was attributed to the different genomic regions that RAPD and AFLP might have used to generate their profiles. Moreover a comparison of AFLP and Inverse Sequence Tagged Repeat (ISTR) in *Agave fourcroydes* revealed genetic variability among clonally propagated plants and suggested that the genetic diversity was generated during the asexual reproduction through rhizomes (Demey et al., 2004; Gonzalez et al., 2003; Infante et al., 2003). The genetic diversity in the agavaceae was also reported by (Gonzalez et al., 2003; Infante et al., 2007) using AFLP. Despite reports of genetic diversity the present state of blue agave genetic research is still confusing because of the contradiction in the results of different genetic fingerprinting techniques. The possible mechanism responsible for the genetic diversity among rhizome derived agave plants is currently unknown. However the genetic variability which has been revealed by ISTR suggests that retrotransposons may be involved in some way in generating this diversity (Infante et al., 2006) .

Retrotransposon based molecular marker systems such as Sequence Specific Amplification Polymorphism (SSAP) and Inter-Retrotransposon Amplified Polymorphism (IRAP) have extensively been used for the evaluation of phylogenetic relationship of many plant species (Pereira et al., 2005; Smykal, 2006; Teo et al., 2005; Vershinin et al., 2003). In many cases the SSAP and IRAP have been more informative than AFLP and other DNA marker technologies for example barley (Waugh et al., 1997), pea (Ellis et al., 1998), tomato (Tam et al., 2005), and cashew (Syed et al., 2005). Recently a retrotransposon based molecular markers SSAP has been

developed and reported to be highly polymorphic in different agave species (Bousious et al., 2007).

1.2 Transposable genetic elements in plants

Transposable elements were first discovered and reported by Barbara McClintock in 1951, while she was studying chromosome breakage in maize and noticed variegation in the colouring of maize kernels (McClintock, 1951). It has been known for some time that eukaryotic genomes have high levels of interspersed repetitive DNA (Britten and Kohne 1968). Large-scale DNA sequencing has revealed that a high amount of the repetitive DNA is derived from the activity of transposable elements. Transposable elements are segments of DNA with the ability to move between different chromosomal locations and replicate within the genome. It has been proven by extensive research in recent years that these mobile elements are ubiquitous components of all eukaryotic and prokaryotic genomes (Miller and Capy, 2004). Transposable elements belong to different classes, each of which uses a particular different replicative strategy, which involves either RNA or DNA intermediates. The broad distribution of all classes of transposable elements across the eukaryotic tree of life indicates that they are long standing residents of eukaryotic genomes (Wicker and Keller, 2007). Transposable elements can be thought of as genomic parasites due to their replicative mode of transposition and life cycle, independent of the host cells functions and replication. Moreover they can replicate faster than the host that carries them which can lead to their high copy number in many genomes (Brookfield, 2005; Kidwell, 2002). Due to their replicative mode of transposition and their maintenance over long evolutionary timescales transposable elements are also referred to as "selfish DNA" (Wicker and Keller, 2007) and their interaction with the host genome is usually described as a host-parasite interaction (Kumar and Bennetzen, 1999). The selfish or parasitic nature as has also greatly influenced the evolutionary trajectory of their hosts due to their insertion modifying the size and function of the host genome (Brookfield, 2005; Deininger, 2003). Transposable elements can be divided into two broad classes on the basis of their sequence organization and transposition mechanism.

1.2.1 DNA transposons (Class II elements)

The transposable genetic elements that transpose via a DNA intermediate are known as DNA transposons or class II elements. During the process of transposition the element is excised from one genomic location and inserted into another one with the help of an enzyme called

transposase while the transposon itself serves as the intermediate. DNA transposons can be divided into three main subclasses, the normal cut and paste transposons that excise as double stranded DNA and reinsert elsewhere (Craig et al 2002), the *Helitrones* which use a mechnism related to rolling circle replication and *Mavericks*, whose replication mechanism is not yet known but they are thought to replicate using a self encoded DNA polymerase (Kapitonov and Jurka, 2006; Pritham et al., 2007). All cut and paste elements are characterised by the presence of transposase gene and terminal inverted repeats (TIRs) (Bennetzen, 2000; Wessler, 2006). *Helitrons* do not have TIRs but rather short terminal conserved motifs and their autonomous copies encode a Rep/Helicase (Kapitonov and Jurka, 2001). *Mavericks* are very long transposons and capable of encoding multiple proteins, most of which are related to double stranded DNA viruses (Feschotte and Pritham, 2005; Kapitonov and Jurka, 2006; Pritham et al., 2007). Class II elements typically are present in fully functional "autonomous" elements which have a transposase gene, but commonly "non-autonomous" elements are also present which lack the complete sequence but can transpose as the missing functions are supplied by the autonomous copies (Kazazian, 2004). The transposition is carried out by the transposase activity which binds to the TIRs as well as target DNA, the element is removed from the excision site by DNA breakage, due to the hydrolysis of phosphodiester bond, and inserted into the new location (Kazazian, 2004). The staggered cuts generated during the excision are repaired by the repair mechanisms of the host genome, terminal site duplications (TSD) of varying lengths typical to each superfamily of the elements are created (Feschotte et al., 2002; Feschotte and Pritham 2007).

Due to the non-replicative mode of transposition DNA transposons are usually in low copy numbers. However the gaps left at the excision sites are sometimes repaired by using the sister chromatid as a template which can result in an increase in the copy number (Bennetzen, 2000). Nevertheless the footprints of a few bases left behind due to imperfect excision can alter the functions of genes and regulatory sequences (Wessler, 2001).

To date at least ten "super families" of cut and paste transposons have been identified in eukaryotes and at least five of these superfamilies are present in plants (Feschotte and Pritham, 2007). These transposon superfamilies have varying lengths of their TIRS and TSDs. The hAT family contains the Dissociator (Ds) and Activator (Ac) elements which were the first DNA elements characterised in maize (McClintock, 1951). On the other hand the Tc1/ mariner family contains Tc1 and mariner like elements and CACTA superfamily contains Suppressor – mutator

(Spm) and dSpm transposons. The other two superfamilies are the mutator superfamily and PIF/ Harbinger family. The length of TIRs in plant DNA transposons varies according to the superfamily of the transposons and ranges from 10bp to more than 1000 bp (Feschotte et al., 2002; Feschotte and Pritham 2007). In recent years the advances in the bioinformatics and availability of sequence data from many organisms has revealed some DNA transposon families which are different from the typical transposon subclasses for example MITEs, the miniature inverted repeat transposable elements and Pack-mules. MITEs are small in size (600bp) but high in numbers and two MITE families, "Tourist" and "Stowaway" account for 3% and 2% of the rice genome respectively (Casacuberta and Santiago, 2003; Feschotte et al., 2002). Another type of elements, the "Pack-MULEs" have also been identified in many plants including maize, *Arabidopsis*, lotus and rice (Casacuberta and Santiago, 2003; Feschotte et al., 2002).

1.2.2 Retrotransposons (Class I elements)

Retrotransposons (class I elements) are different from class II elements in that they replicate via a RNA intermediate. In this system an RNA transcript is produced from the retrotransposon DNA which is inserted in the host genome. The transcript (RNA genome) is produced which is then reverse transcribed to produce a complementary DNA (cDNA) copy which is inserted into the host genome at a new genomic location. Due to their replicative mode of transposition, which leaves the original copy intact, retrotransposons often have amplified themselves to very high copy numbers in many species with plants containing particularly high numbers (Kumar and Bennetzen, 1999). On the basis on the comparison of their genome structures retroelements can broadly be divided into retrotransposons, retroviruses and retroposons (Eickbush and Jamburuthugoda, 2008; Peterson-Burch and Voytas, 2002). Retrotransposons are similar to retroviruses in their replication mechanism and genome organisation while retroposons are very similar except they lack their own reverse transcriptase. Retrotransposons can be further subdivided into "LTR retrotransposons" and "non LTR retrotransposons". LTR retrotransposons are the most abundant class of retrotransposons (Kumar and Bennetzen, 1999; SanMiguel et al., 1996) which are comprised of two major families of elements the Pseudoviridae and Metaviridae (Boeke et al., 2006). The Pseudoviridae contains Pseudovirus, Hemivirus and Sirevirus while Metaviridae contains Metavirus, Errantivirus and Semotivirus (Boeke et al., 2006; Eickbush and Jamburuthugoda, 2008) . The retrotransposons in Pseudoviridae are synonymous with the Ty1-*copia* group whilst the Metaviridae are synonymous with the Ty3-*gypsy* group retrotransposons.

The detailed description of LTR retrotransposons is explained further in the later parts of this chapter.

1.2.3 Non-LTR retrotransposons

Non-LTR retrotransposons were first discovered in mammalian genomes but have also been identified in plants, fungi and invertebrates. In general they can be distinguished as LINEs (long interspersed nuclear elements) and SINEs (short interspersed nuclear elements) (Schmidt, 1999). LINEs are autonomous elements while SINEs have no coding capacity. LINEs can be several kilobases long and transpose autonomously by encoding gag like protein, reverse transcriptase and endonuclease. However the absence of LTRs clearly separates them from LTR retrotransposons and retroviruses. In contrast to the LTR retrotransposons the genome of LINEs contains an untranslated region (UTR) at 5' end and the 3' end terminates in a polyadenylated (polyA) tail. There are two open reading frames in LINEs, ORF1 encodes the gag like protein while ORF2 encodes a protein with reverse transcriptase and endonuclease functions (Schmidt, 1999). LINEs are transcribed into RNA by an RNA polymerase II, the RNA is translated into a ribonucleoprotein complex in the cytoplasm, and integrated into new genomic location through a mechanism called target-primed reverse transcription (TPRT) (Schmidt, 1999; Yang and Kazazian, 2006). LINEs are especially abundant in mammalian genomes, for example L1 retrotransposon occupies about 17% of the human genome (Yang and Kazazian, 2006). Only a few LINEs have been studied in plants but the number is gradually increasing. The genomes of *Lilium* species contain a full length LINE called *del2* (Leeton and Smyth, 1993) while they have also been reported in the genome of barley with the name BLIN (Schmidt, 1999). LINEs are also a part of some organelle genomes such as the mitochondrial genome of *Arabidopsis thaliana* (Knoop et al., 1996).

SINEs are also called retroposons, they are very small and nonautonomous in nature as they have no coding domains in their genome (Schmidt, 1999). SINEs have been extensively studied in mammals and are reported to be of high copy number, for example the 300bp *Alu* element is small but very abundant in the human genome with more than a million copies comprising 10% of the genome (Batzer and Deininger, 2002). Most SINEs are derivatives of cellular non-viral RNA and are composed of three major regions: a 5' tRNA related region, an internal tRNA unrelated region and a 3' AT rich region (Shedlock and Okada, 2000). As SINEs do not encode any proteins themselves it is proposed that they use the enzymatic machinery of LINEs for their

retrotransposition (Oshima et al., 1996; Sun et al., 2007). They are ubiquitous in eukaryotic genomes including plants (Sun et al., 2007; Zhang and Wessler, 2005). However only a few SINE families have been identified in plants for example *p*-SINE*1* family of SINEs in rice (Moochizuki et al., 1992), S1 retroposons of brassica (Deragon et al., 1996) and Ts family of elements in tobacco (Yoshioka et al 1993).

1.2.4 Genome organization and replication strategy of LTR retrotransposons

LTR retrotransposons were named because of the presence of the long terminal repeats which flank them. They are the most abundant and wide spread type of class I transposable elements (Casacuberta and Santiago, 2003; Kumar and Bennetzen, 1999). LTR retrotransposons are a major component of eukaryotic genomes but their proportion is different in different organisms. They can comprise large proportions of some plant genomes in contrast they are not the major type of retrotransposons in mammalian genomes (Vitte and Bennetzen, 2006). LTR retrotransposons are considered to be the major driving force responsible for the nuclear genome expansion in plants and the differences in the genome size of different plant species is attributed to them. They can comprise 80-90% of the total DNA content of some large genomes but they are underrepresented in some small plant genomes, for example they comprise 17% of the rice genome (McCarthy et al., 2002) and only 5.6% of the genome of *Arabidopsis thaliana* (Pereira, 2004).

The two main LTR retrotransposon families, the Pseudoviridae and Metaviridae contain the same genes; however they are arranged in a different order in each family. The gene order for Ty1-*copia* elements is GAG-PR-RT-IN- RNH, while in Ty3-*gypsy* elements the gene order is GAG-PR-RT-RNH-IN (Eickbush and Jamburuthugoda, 2008; Kumar and Bennetzen, 1999) .

The gag gene encodes a short structural nucleocapsid protein, which promotes the formation of a virus-like particle (VLP) which packages the single stranded RNA and the other proteins before they are transferred to the cytoplasmic phase of the life cycle, where reverse transcription takes place within the VLP (Peterson-Burch and Voytas, 2002). The products of other four genes are collectively called the pol polyprotein that contains enzymic function required for the replication. The pol gene comprises of protease (PR), which is responsible for gag processing and maturation other proteins from the polyprotein precursor, (RT) reverse transcribes the RNA into double stranded cDNA and the integrase (IN) mediates insertion of the extrachromosomal double stranded cDNA copy in the new genomic location. The N-terminus is conserved among most

LTR retroelements and retroviruses, whilst the C-terminus is highly variable even between members of the same family (Malik and Eickbush, 1999; Peterson-Burch and Voytas, 2002). IN performs the cutting and joining reactions by interacting with the cDNA as well as the target site. It processes the blunt ends of the linear cDNA copy to produce 3' hydroxyl ends thus creating a staggered cleavage at the host site that results in the distinctive terminal site duplication (TSD) after insertion (Katz and Skalka, 1994). The IN is followed by reverse transcriptase gene (RT) which is the best studied and highly conserved domain of retroelements with (Xiong and Eickbush, 1990). RT and the associated RNaseH gene that lies at the 3' end of the pol polyprotein are responsible for the processing of the single stranded RNA template into a linear double stranded cDNA within the VLP. RT is a multifunctional enzyme with RNA-dependent and DNA-dependent DNA polymerase abilities. However reverse transcription is an error prone process due to the lack of proofreading and repair mechanisms. The RNaseH enzyme coordinates with the RT in reverse transcription (Katz and Skalka, 1994).

LTR retrotransposons are flanked by non-coding direct terminal repeats that vary in size between 0.1-5kb which contain the *cis*-acting regulatory motifs and the promoter and polyadenylation signals required for the initiation and termination of transcription. LTRs are divided in 3 distinct regions in U3-R-U5 order. The U3 contains the transcriptional regulators, whilst the promoter box at its 5' end and termination signal at its 3' are the characteristics of the R region (Kumar and Bennetzen, 1999). A primer binding site (PBS) links the 5'LTR with gag gene and acts as a target site for tRNA molecule which primes the synthesis of first strand of the cDNA molecule. Similarly a linker domain separates the 3'LTR from the pol gene and a poly purine tract PPT at the 5' end of the linker domain serves as a primer for the synthesis of second strand of cDNA (Wilhelm and Wilhelm, 2001). The transcribed RNA molecule is translated in the cytoplasm by the formation of VLPs and the synthesis of cytoplasmic double stranded cDNA intermediate (Wilhelm and Wilhelm, 2001). The life cycle and genomic organization of retroviruses is the same as LTR retrotransposons, however they contain an envelope (env) gene in the linker domain which allows the infection of retroviruses to another cell.

1.2.5 Contribution of LTR retrotransposons to the structure of plant genomes

A major factor contributing to the large difference in the genome size of plants, and especially the angiosperms is the repetitive DNA (Bennett and Leitch, 2005; Flavell et al., 1974). Although polyploidy and chromosome duplication are also responsible for large differences in genome

size, one of the most significant contributors to genome size variation are the LTR retrotransposons (SanMiguel et al., 1996; Vicient et al., 1999a). Due to the replicative mode of transposition of retrotransposons it was initially proposed that retrotransposons are only able to *increase* genome size (Bennetzen and Kellogg, 1997).

The role of LTR retrotransposons in influencing genome size has been studied in a wide variety of plants but the main focus has been the economically important grass species. LTR retrotransposons occupy 17% (McCarthy et al., 2002; Sanmiguel and Bennetzen, 1998) of rice, 50-80% of maize (Meyers et al., 2001) and 70% of barley genomes (Vicient et al., 1999b). Large "seas" of LTR retrotransposons surround small "islands" of genes. However the differences in the intergene LTR retrotransposons in plants from the same family suggests that LTR retrotransposons are the driving force responsible for the increase in genome size of grass species (Bennetzen, 2000). Dense gene regions surrounded by nested clusters of Ty1-*copia* and Ty3–*gypsy* retrotransposons have been found in the 3.8Mb of available wheat DNA sequence (Sabot et al., 2005) and the same pattern was observed in the 66kb Rar1 gene locus in barley (Shirasu et al., 2000) while nested clustering of LTR retrotransposons in the region of maize genome flanking adh1 gene was also reported (SanMiguel et al., 1996). A comparison between the adh1 regions of maize and sorghum revealed that the adh1 region of sorghum is devoid of retrotransposons, which suggests that these LTR retrotransposons were inserted in the maize genome after the divergence of maize and sorghum about 2-6 Mya (Sanmiguel and Bennetzen, 1998). This suggests that LTR retrotransposons are a major contributor to the differences in the genome size of maize and sorghum. The genome size of *Oryza australiensis* is double the size of its cultivated relative *Oryza sativa* due to the activity of three LTR retrotransposons namely RIRE1, Kangourou and Wallabi which collectively make up 60% of the *Oryza australiensis* genome (Piegu et al., 2006).

It has been suggested that retrotransposons undergo short waves of activity followed by varying periods of silence (Wicker and Keller, 2007). Instead of the increase only model it is now suggested that the genome evolution and genome size differences should result from the balanced forces: increase induced by retrotransposition and decrease caused by counterbalancing forces like recombination and deletions (Petrove, 2002: Vitte and Panaud, 2005).

1.2.6 Counterbalancing mechanisms of LTR retrotransposon-mediated genome expansion

The accumulation of LTR retrotransposons in a few million years have played a vital role in shaping the architecture of plant genomes and this must have been the result of a large amount of retroelement activity. In apparent contradiction of this observation very few active elements have been detected (Feschotte et al., 2002). Most retrotransposons are usually silent but can be activated under stress conditions (Grandbastien, 1998). It has been shown that activity of retroelements is regulated at transcriptional level (Melayah et al., 2001) so host genomes have developed epigenetic mechanisms for the repression of retrotransposons which is probably to limit the amount of damage that these elements can inflict on the genome (Feschotte et al., 2002). Research in recent years has suggested an increase-decrease mechanism for the genomes of plants with respect to their retrotransposon content (Bennetzen et al., 2005). A balance between retrotransposon insertion and removal through recombination and deletion has been suggested to be responsible for the evolution of genome size in plants (Petrove, 2002: Vitte and Panaud, 2005). Two types of removal mechanisms have been reported to occur in plant genomes, unequal homologous recombination and illegitimate recombination (Bennetzen et al., 2005). Unequal homologous recombination removes the LTR retrotransposons through solo LTR formation and thought to be the major removal mechanism in *Arabidopsis* (Devos et al., 2002; Vitte and Panaud, 2003). By the evaluation of retrotransposon removal from the barley it was suggested that solo-LTR formation could be the possible mechanism responsible for the contraction of genome size since the recombination between the LTRs of two elements can delete a whole copy of an element and contract massive amplifications of retrotransposons (Shirasu et al., 2000). Unequal homologous recombination has also been reported in rice through analysis of three gypsy like LTR retrotransposon families (Vitte and Panaud, 2003). However illegitimate recombination could also be responsible for the genome size reduction by causing small deletions in *Arabidopsis thaliana* (Devos et al 2002). Deletions have also been suggested to be responsible for the genome size differences in *Drosophila* (Petrove and Hartle, 1998). Collectively the contraction of retrotransposon amplification through unequal homologous recombination and the formation of deletions through illegitimate recombination could be responsible for the differences in the genome size of different plant species (Vitte and Panaud, 2005).

1.2.7 LTR retrotransposons as mutagens

The movement of transposable elements particularly LTR retrotransposon can generate a wide variety of mutations in plant genomes. The first ever plant retrotransposon, *Bs1* was characterised following its insertional inactivation of the maize *adh* gene (Johns et al., 1985). The isolation of the first active plant retrotransposon, *Tnt1* was also a result of its insertion within the *Nitrate reductase* gene of tobacco (Grandbastien et al., 1989). Many examples of mutations due to insertional mutation caused by retrotransposons have been reported (Vignols et al., 1995; Takano et al., 2001). However mutation can also be caused by the insertion of retrotransposons in non coding regions. The insertion of retrotransposons within introns can result in tissue specific alternative splicing leading to the production of fully active or truncated proteins in different tissues (Marillonnet and Wessler, 1997; Leprince et al., 2001; Varagona et al., 1992). LTR retrotransposons can also generate mutations by inserting in the non-coding regions close to genes (Hirochika, 2001). Retrotransposons that insert within or near genes are usually the members of less abundant families. On the other hand higher copy number retrotransposons tend to insert in nongenic and heterochromatic regions thus creating nested clusters between genes without causing mutations (Feschotte et al., 2002).

Besides insertions within genes or *cis*-acting regulatory sequences that can alter gene function, retrotransposons can interact with their host genomes in many ways that can equally be useful to them. The host genome can eliminate the harmful effects of retrotransposon and can take advantages from the changes that can positively influence the host genes (Kumar and Bennetzen, 1999). Although retrotransposons can modify the expression and coding capacity of genes and transposition can be an extremely deleterious event and sometimes they may not give rise to genome remodelling mechanisms as proposed by Barbara McClintock (McClintock, 1984) but many evidences have shown that a large number of retrotransposons are activated by stresses and their mobility has shaped the eukaryotic genome on many ways (Casacuberta and Santiago, 2003).

1.2.8 The Evolutionary relationships of the different retrotransposon families.

The transcription and translation mechanisms of non–LTR retrotransposons indicates the prokaryotic origin of non-LTR retrotransposons (Malik, 2005). It has been suggested that early non-LTR retrotransposons did not have their own RNaseH domain and they may have relied on

the host encoded RNaseH activity for their transcription and translation. In contrast all LTR retrotransposons have a functional RNaseH gene, which indicates that non-LTR retrotransposons might predate the LTR retrotransposons (Malik, 2005). The RNaseH activity of eukaryotic host cells was restricted to the nucleus and organelles while LTR retrotransposons use a replication strategy which is completed in two steps, which take place in nucleus and cytoplasm. It has been proposed that LTR retroelements originated in the early eukaryotic cell from a fusion of a DNA transposon carrying an integrase-like domain and non-LTR element with reverse transcriptase domain (RT) (Malik and Eickbush, 2001)

The phylogenetic analysis of reverse transcriptase sequences from Ty1-*copia* and Ty3-*gypsy* elements suggests that these two families of retrotransposons shared a common ancestor (Boeke et al., 2006). Moreover the life cycle as well as the structure and functions of genes encoded by LTR retrotransposon is similar to that of retroviruses (Eickbush and Jamburuthugoda, 2008; Malik and Eickbush, 1999). It is suggested that LTR retrotransposons might have resulted from the defective retroviruses which had lost their ability to infect and move across different cells. Consequently they might have restricted within the cell with the ability of reverse transcription and transposition. It is also possible that an LTR element might have acquired an envelope gene with all functions for invasion across the cell membrane, in the course of its evolution thus giving rise to retroviruses (Malik and Eickbush, 1999). Phylogenetic relationships between Ty3-*gypsy* elements and vertebrate retroviruses and sequence similarities between the envelope gene of insects and viruses have also been found (Malik and Eickbush, 1999; Malik et al., 2000). It is suggested that *envelope* glycoproteins associate with the cell membrane and facilitate the budding of viral core particles from infected cells. They also mediate infection by recognizing cellular reporters(Peterson-Burch et al., 2000). However plant cells contain the cell wall which acts as a barrier for transmembrane transmission of viruses, thus challenging the role of the *envelope* gene in plant retrotransposons. However it could be imagined that they might have lost the *envelope* function (Bennetzen, 2000) or might have transformed its function to a new recalibrated intracellular role.

1.3 Retrotransposon expression, stress activation and somaclonal variations

The abundance of retroelements particularly LTR retrotransposons in many species is a consequence of their replicative mode of transposition. Retroelements employ a ‘copy and paste’ mechanism, which leads to element accumulation and host genome expansion, as the parental

copy as well as the descendants are stable insertions capable of initiating another round of replication (Kumar and Bennetzen, 1999). However, the life cycle and genome organization of each retroelement type is considerably diverse and varies in complexity, reflecting the different strategies developed in order to cope with the increasing evolutionary demands of the eukaryotic cell.

The expression of retrotransposons in plants and other eukaryotic organisms is regulated, thereby regulating the transposition frequency in the host genome. The evolution of controlled mechanisms for the transcription and transposition of retrotransposons in the host genome may be crucial to minimise their possible deleterious effects on the host (Kumar and Bennetzen, 1999). Retrotransposons transpose via an RNA intermediate and without the availability of an RNA template for reverse transcription, the transposition would not be possible. So the simplest way to control their activity would be via regulation of transcriptional initiation (Kumar and Bennetzen, 1999). A correlation of transcription and transposition has been demonstrated for Tt01 of tobacco (Hirochika, 1993) and *Tos17* retrotransposons of rice (Hirochika et al., 1996a). For example the expression of *Tto1* and *Tos17* is associated with an increase in the level of their RNAs, pointing toward the fact that transposition of these elements is regulated mainly at transcriptional level. An increase in the copy number of *Tto1* associated with an abundance of transcripts was observed in the cultural cells (Hirochika, 1993) and an increase in the copy number of *Tos17* was also seen in the tissue culture (Hirochika et al., 1996a). However transposition is not seem to be regulated by transcription in other elements, for example in spite of high level of transcripts in the leaves, transposition of BARE-1 has not been observed in barley (Suoniemi et al., 1996b) which means that although transcription is a pre requisite for the transposition of retrotransposons but other factors are also required and may be regulated (Kumar and Bennetzen, 1999).

In animals and yeast the expression of retrotransposons is under the effect of hormonal and developmental factors but a general picture of expression of many plant retrotransposons is difficult to establish because of the lack of comparative studies in different plant tissues (Grandbastien, 1998). Most plant retrotransposons are inactive or silent in somatic tissues but are expressed during certain stages of plant development, for example Tnt1 is only expressed in roots and at very low levels (Pouteau et al., 1991) while Tto1, Tto10 and Tos17 are not expressed in leaf tissues (Hirochika, 1993; Hirochika et al., 1996a). The expression of PREM-2 elements of maize was detected in early microspores and expression of Opie, Huck, Cinful (Avranmova et

al., 1995; (SanMiguel et al., 1996) and BARE-1 (Suoniemi et al., 1996a) was detected in the leaf tissues.

Like all other expressed sequences in eukaryotes, retrotransposons are differentially active in different tissues, at different stages of development and under different regimes (Bennetzen, 2000). Whether viewed as parasitic or selfish DNAs or as mobile elements with some beneficial role, it is clear that a very high level of activity of these elements can be deleterious to the individuals presumably due to genic and chromosomal mutations (Bennetzen, 2000). Due to this reason the eukaryotic genomes seem to have developed mechanisms to reduce the activity of transposable elements and control their expression and mutagenic activity (Casacuberta and Santiago, 2003). There are several mechanisms which can be responsible for the inactivation of retrotransposons in the eukaryotic genomes; among them silencing mechanisms are probably the most general and effective. Post-transcriptional gene silencing (PTGS) is a sequence specific RNA degradation that probably constitutes a general antiviral defence mechanism in plants. PTGS operates through the production of small RNAs (Bernstein et al., 2001; Matzke et al., 2001) and the identification of small RNAs of 21-22 nucleotides derived from tobacco LTR retrotransposon suggests the presence of PTGS as regulatory mechanism for plant LTR retrotransposons (Okamoto and Hirochika, 2001).

Transcriptional gene silencing (TGS) is another silencing mechanism which acts as a promoter inactivation mechanism possibly directed to abolish the transcription of mobile elements (Vance and Vaucheret, 2002; Vaucheret and Fagard, 2001). TGS is influenced by different factors but presence of multiple copies of the target sequence seems to be a major factor leading to the gene silencing. For example, the activity of Drosophila I elements is repressed by the introduction of multiple copies of a transgene expressing a small internal region of this element (Jensen et al., 1999) and the tobacco retrotransposon Tto1 becomes silent in *Arabidopsis* due to an increase in its copy number (Hirochika et al., 2000). The presence of short interfering RNA (siRNA) corresponding to retrotransposons sequences (Hamilton et al., 2002; Lalve et al., 2002) also confirms that these elements targeted by the silencing mechanisms like TGS (Casacuberta and Santiago, 2003). In spite of the presence of silencing mechanism some retrotransposons have maintained their ability to transpose particularly under stress conditions (Grandbastien, 1998). There can be many reasons of the association of transposon mobility with the stress but it would well be that the silencing mechanisms are somehow relaxed under these situations, allowing retrotransposons to temporally escape the genomic control (Casacuberta and Santiago, 2003).

Although most retrotransposons are inactive or silent in somatic tissues of plants but active in response to the stressful conditions (Grandbastien, 1998) some biotic and abiotic stresses can activate the retrotransposons hence increasing the level of transcripts in plants, for example Tnt1 (Beguiristain et al., 2001; Pouteau et al., 1994; Pouteau et al., 1991), Tto1 (Hirochika et al., 1996b; Takeda et al., 1998; Takeda et al., 1999) Tto2 (Hirochika et al., 1996a) of tobacco and Tos17 (Hirochika et al., 1996a) of rice. The activation of tobacco retrotransposons Tnt1 and Tto1 is greatly increased by various abiotic stresses, including protoplast isolation, cell culture, wounding, methyl jasmonate, $CuCl_2$, and salicylic acid similarly various biotic stress factors like fungal extracts or inoculation with bacterial, viral and fungal pathogens can activate the transcription of retrotransposons (reviewed by Kumar and Bennetzen, 1999).

The ability of plant LTR retrotransposons to transpose in response to stressful conditions appears to be linked to the *cis* regulatory elements in their promoter (U3 region of LTR). These *cis* regulated elements are associated with the signal transduction pathways related to plant defence response. For instance, the 5´ LTR of the BARE1 retrotransposon contains ABRE elements that respond to ABA (Suoniemi et al., 1996a). In the case of Tto1 a13 bp motif has been identified as a cis regulatory sequence associated to its induction by jasmonic acid (Takeda et al., 1999). Interestingly, this motif specifically binds different transcriptional factors, one of which, that has been named as LBM1, is identical to previously identified MYB-1 factor induced by virus infection (Sugimoto et al., 2000) while over-expression of another MYB (ntMYB2) transcription factor activates both Tto1 and the PAL defence related gene in tobacco (Sugimoto et al., 2000). Three different subfamilies of Tnt1, Tnt1A, Tnt1B, and Tnt1C (Vernhettes et al., 1998) are transcribed under stress situations associated to the plant defence reactions (Beguiristain et al., 2001; Casacuberta et al., 1997).

The promoter of *Tnt1*A contains two different boxes located in the U3 region of the LTR which is important for the element's response to JA and has sequence similarities with plant defence promoter (Vernhettes et al., 1997). On the other hand Tnt1B and Tnt1C are also expressed in tobacco under different stress situations. *Tnt1*C can be induced by the treatment of the leaves with salicylic acid as it contains as as-1 element in its U3 region required for the response to salicylic acid and auxin, while *Tnt*1B is expressed in tissue culture (Beguiristain et al., 2001). The promoter of *TLC1.1* retrotransposon from tomato is activated by multiple stress related signals. This promoter integrates different signal transduction pathways that regulate plant developmental processes and adaptation to environmental cues. A particular array of *cis*

regulatory elements seems to be responsible for the capability of this promoter to be induced by different signal molecules (Salazar et al., 2007). It is well established that retrotransposons are activated in defence related stresses because their promoters are similar to plant defence genes and bind the same defence induced factors (Casacuberta and Santiago, 2003). Retrotransposons are structurally and functionally very similar to retroviruses and has been proposed that, as retroviruses, they could display a high sequence plasticity allowing them to rapidly evolve parts of their sequence, and acquire stress associated promoters (Casacuberta et al., 1997). Due to this possibility a high variability of Tnt1 U3 region has allowed this family of retrotransposons to evolve three different stress inducible promoters in tobacco (Beguiristain et al., 2001) and another Tnt1 related element *Retrolycl* has evolved different promoters in tomato (Araujo et al., 2001). Although many retrotransposons are turned on by biotic and abiotic stresses but stress is a rare event and thus stress induced retrotransposons will transpose in away that the host genome viability is not compromised. On the other hand the variability generated by the movements of these elements could help the genomes to rapidly evolve when facing a situation to which they are not well adapted, as it was initially proposed by McClintock (McClintock, 1984).

It is well known that many retrotransposons are activated under the effect of biotic and abiotic stresses including tissue culture (Grandbastien, 1998). Several plant retrotransposons have been shown to be activated under tissue culture (Kaeppler et al., 2000). As plant tissue culture is a method by which plant clones are obtained and the plants regenerated from tissue culture it could be assumed that the cells from individual clones are genetically identical. This however is not always the case and a number of genotypic instabilities have been observed to occur in tissue culture derived plants which are at least partly due to *in-vitro* induced stress (Evans et al., 1984; Larkin and Scowcroft 1981). This phenomenon of variation among cultured cells and plants derived from them is called somaclonal variation (Larkin and Scowcroft 1981). The molecular basis of somaclonal variation is not precisely known but genetic and epigenetic mechanisms have been proposed to be responsible.

Somaclonal variation may arise as a result of point mutations, rearrangements of nuclear or organelles DNA, polyploidy and epigenetic changes causing deviations from the desirable phenotype quality standards (Phillips et al., 1994; Jaligot et al., 2000) but stress activation of plant retrotransposons is thought to be one of the major causes of somaclonal variations (Alves et al., 2005). In several studies of morphologically abnormal *in vitro* regenerants of plants no changes in the nucleotide sequences were detected, therefore focus turned towards epigenetic

factors (Martienssen and Colot 2001). However due to the advances in retrotransposon research in the last few years and in recognition of the fact that retrotransposons are activated by biotic and abiotic stresses including tissue culture (Grandbastien, 1998; Wessler, 1996) many attempts have been made to understand the stress activation of retrotransposons (Grandbastien et al., 1997; Hirochika, 1993; Hirochika et al., 1996a) and their possible link to the somaclonal variation in plants (Smykal et al., 2007; Alves et al., 2005; Li et al., 2007).

In the work described here a population of Ty1-*copia* retrotransposons have been isolated from *Agave tequilana.* The copy number of these has been investigated and estimated in the *A. tequilana* genome and the activity and insertional polymorphisms of these elements has been studied during vegetative propagation and in *A. tequilana* tissue culture lines using retrotransposon based molecular markers.

1.4 Aims of the thesis

The aims of this work were:

i) to isolate and characterise the Ty1-*copia* retrotransposon population of the blue agave (*Agave tequilana*).

ii) to evaluate and estimate the abundance of Ty1-*copia* retrotransposons in the genome of *Agave tequilana.*

iii) to study the activation expression and insertional polymorphism of the above retrotransposons in response to tissue culture and vegetative propagation.

CHAPTER 2

General Materials and Methods

2.1 Plant Material

The plant material used was the blue agave (Weber, variety azul) from the Sussex Botanical Collection, University of Sussex, UK .

2.2 Nucleic acid Isolation and Purification

2.2.1Plant DNA Extraction

Plant DNA from agave leaves and tissue culture was isolated using the DNeasy Plant Mini Kit (QIAGEN Cat.69104, GmBH Germany) according to the manufacturer's instructions. The pestles and mortars were washed and autoclaved every time before they were used for grinding of the plant material. The plant leaf and tissue culture material was ground in liquid nitrogen before using the DNeasy Plant DNA extraction protocol. The kit was used through out this study for the extraction of plant DNA.

2.2.2Total RNA extraction

Total RNA was extracted using RNeasy Plant mini kit by QIAGEN cat no: 74904 in accordance with the manufacturer's instructions.

2.2.3mRNA isolation

mRNA was purified using QIAGEN Oligotex® mRNA Mini Kit cat no:70022. The starting material for the purification of mRNA was total RNA and extractions were carried out in accordance with the manufacturer's instructions.

2.2.4 cDNA Synthesis

The cDNA was synthesized from the mRNA using QIAGEN Quanti Tect® cat no: 205311.

2.2.5 Plasmid DNA isolation

Plasmid DNA from the bacterial colonies was isolated by alkaline lyses method. The selected colonies were grown overnight at 37°C in 2ml LB liquid medium with 2µl of ampicillin (50mg/ml) with shaking at 80 rpm.

1ml of the liquid culture was transferred into a 1.5ml eppendorf tube, centrifuged at 13000 rpm for 1 minute and the supernatant was discarded. The pellet formed was resuspended in 100 µl of solution I (50mM Glucose, 25mM Tris.HCl pH 8, 10mM EDTA pH 8). 200 µl solution II (0.2N NaOH, 1%SDS) was added and mixed by inverting the tube following an incubation at room temperature for 5 minutes.The sample was incubated on ice for 5 minutes after adding 150 µl solution III (5MKOAc 60ml,Glacial acetic acid 11.5 ml,water 28.5 ml) following a spin for 5 minutes at 13000 rpm. The supernatant was discarded and the precipitate was cleaned twice with 100 µl of phenol following a spin at 13000 rpm. The plasmid DNA was precipitated by centrifuging the pellet in 1ml of 100% ethanol at 13000 rpm and discarding the supernatant. The pellet was then washed twice in 100 µl of 70% ethanol following a centrifugation for 2 minutes. The pellet was dried in a heating block for 30 minutes at 30°C and resuspended in 402µl of 1xTE buffer. Finally 1µl of RNase A was added and the tube was incubated for 1 hour at 37°C. The samples were visualized by running a 1% agarose gel. In the later stages of this study the plasmid DNA was isolated using QIAprep® Spin Miniprep Kit catalog no: 27104 by QIAGEN, GmbH, Germany in accordance with the manufacturer's instructions.

2.2.6Purification of DNA fragments from agarose gel

The PCR products were resolved on a 1% agarose gel and the fragments of interest were separated from the gel. These fragments were recovered by gel purification using QIAQUICK Gel Purification Kit by QIAGEN, GmBH Germany.

2.2.7 PCR Purification

Purification of PCR products was carried out using QIAQUICK PCR Purification Kit by QIAGEN, GmBH Germany cat: 28104. Purification of the PCR products was performed to

remove unincorporated primers and reagents as these will cause problems with the subsequent sequencing reactions.

2.3 Gel Electrophoresis

2.3.1 Agarose gel Electrophoresis

1% agarose gel was made by dissolving 1g of agarose in 100ml of 1x TBE buffer (0.08M Tris base, 0.08 M boric acid, 0.020M EDTA) and heating in a microwave oven. 4 µl of ethidium bromide (10mg/ml) was added to the gel. DNA samples were mixed with 5 x loading buffer (0.2% bromophenol blue and 30% glycerol in TBE buffer) were loaded in the wells of the gel and run for 1hour at 80V in 1X TBE buffer. The gel was visualized on a UV trans-illuminator and photographs of the gel were taken by a camera which was fitted in a chamber containing UV trans-illuminator.

2.3.2 High resolution agarose Gel Electrophoresis

2% high resolution agarose gel was made as above with Agarose Extra by Alpha Labs. UK. The electrophoresis was carried out at 85 V for 8 hours, visualized as above and photograph was taken.

2.3.3 Polyacrylamide Gel Electrophoresis

To make polyacrylamide gel 80ml of polyacrylamide gel stock solution (306g urea, 66.6ml of 5xTBE, 40% acrylamide, total volume upto 666ml with water) was degassed using vacuum degasser, 90 µl of 25% ammonium persulphate and 90 µl of TEMED was added and the solution was gently mixed to avoid the formation of bubbles.

The glass plates used to prepare the gel were washed with distilled water, cleaned and dried with 70% and 100% ethanol respectively. The plates were prepared by placing spacers (width 0.4 mm) between them and metal clips to hold them together. A piece of Watman paper was used in the bottom end of the gel chamber to stop the leakage of the gel. The gel solution was poured between the plates with constant tapping to avoid the formation of bubbles. An inverted comb was placed on top of the gel between the plates to create a space for loading. This gel was left for

at least an hour to set before placing it into the gel rig. The upper and lower chambers of the apparatus were filled with 0.5x TBE buffer, the comb was placed between the plates pointing downwards making wells for the loading. Before loading on the gel an equal volume of loading dye (9.5ml formamide, 200µl 1MEDTA, distilled water 300 µl, 0.5mg bromophenol blue) was added to the samples .The samples were denatured at 94°C for 5 minutes and immediately placed on ice for 5 minutes. The wells were loaded with 4µl of the samples and the gel was run in a vertical electrophoresis apparatus at 80 W until the samples reached the end of the gel.

After the completion of electrophoresis the gel was carefully separated from the plates and fixation was carried out in the fixative solution (10% methanol, 10% acetic acid) for at least 30 minutes. The gel was then transferred to 3mm Whatman chromatography paper, covered with cling film and dried on a vacuum drier for 1 hour. The dried gel was then placed in an autoradiograph cassette and exposed to X-ray film overnight. The film was developed in *Konica SRX 101A* developer.

2.3.4 Formaldehyde gel Electrophoresis

Total RNA was resolved on a 1.4% Formaldehyde gel (1.4g agarose,18ml formaldehyde, 1ml 1M NaH_2PO_4, 81ml RNase free water) at 70V for 1 hour with an equal volume of loading buffer (4µl 1M NaH_2PO_4, 35µl formaldehyde, 100µl formamide, 0.5mg bromophenol blue). The agarose was dissolved in water by heating in a microwave oven. The dissolved agarose was cooled down to about 55°C before adding formaldehyde and NaH_2PO_4. The gel was made in a fume hood to avoid fumes of formaldehyde. The electrophoresis was carried out in 10mM sodium phosphate buffer.

2.4 Polymerase Chain Reaction (PCR)

The polymerase chain reaction (PCR) is a widely used technique in molecular biology. Its name comes from the DNA polymerase used to amplify a fragment of DNA by replication. As PCR progresses, the DNA generated is used as a template for replication. In this way by a chain reaction the DNA template is exponentially amplified. The PCR can amplify a single or a few copies of a fragment of DNA across several orders of magnitude, generating millions or more

copies of the DNA fragment. The conditions of PCR can be extensively modified for particular uses.

2.4.1 Components of PCR

A basic PCR set up requires several components and reagents. These components include:

DNA template: DNA template is the DNA used in the reaction which is to be amplified.

Primers: Normally two synthetic oligonucleotides are used which prime the synthesis by annealing to the template DNA.

DNA polymerase: A temperature stable DNA polymerase (Taq DNA polymerase).

Deoxynucleoside triphosphates (dNTPs): These are the building blocks from which the DNA polymerase synthesizes a new DNA strand.

Buffer solution: The buffer provides a suitable chemical environment for the activity and stability of the DNA polymerase.

2.4.2 Procedure of the PCR

The PCR consists of a series of cycles, each cycle consisting of 2-3 temperature steps. A standard PCR includes:

Denaturation*:* Melting of DNA template at high temperature (94-98°C) by disrupting the hydrogen bonding between complimentary bases of two strands of DNA, generating single stranded DNA.

Annealing*:* The reaction temperature is lowered (45-65°C) to allow the annealing of primers to the single stranded DNA. The annealing temperature is usually the same or lowers than the lowest Tm primer.

Elongation*:* Temperature of this step 72°C which is optimum for *Taq* DNA polymerase to synthesizes a new DNA strand complimentary to the DNA template in 5´ and 3´ direction.

Cycling*:* PCR works in cycles. The fragment of DNA which is denatured and elongated is subjected to another round of denaturation and so on. Specific numbers of cycles can be used according to the requirement of reaction.

Final extension*:* This step is performed at a temperature of 70-74°C but usually 72°C for 5-15 minutes to ensure that all single stranded DNA is fully elongated.

Final Hold*:* In this step usually a temperature of 4°C is employed for an indefinite time and could be used for the short term storage of the reaction.

Throughout the course of this study PCR was used as a key molecular biology technique. A standard 50µl PCR reaction was normally used with 5µl of 10 x PCR buffer with 15 mM $MgCl_2$,

8µl of dNTPs (1.25mM each), 0.4µl *Taq* DNA polymerase (5U/µl New England BioLabs), 1µl of DNA template, 1µl of each of the two primers (0.15 µg/µl), and 33.6µl of sterilized distilled water. The reaction was carried out at 94°C for 1minute, 50°C for 1 minute, 72°C for 1 minute following a final extension at 72°C for 7 minutes.

Changes were made to the standard PCR reaction according to the requirements of individual experiments.

2.5 Subcloning of PCR products and Sequencing

2.5.1 Subcloning of PCR products

The subcloning of the PCR products was carried out using TOPO™ TA Cloning® (Invitrogen). The technology takes advantage of the terminal transferase activity of *Taq* polymerase, which is non-template dependent, and adds a single overhanging A to the 3'end of the PCR products. The TOPO™ TA cloning vector has a single overhanging T at its 3'ends to facilitate the efficient ligation of the PCR product. The vector is also 'activated' by a covalently bound topoisomerase I, which cleaves the phosphodiester backbone of the DNA after the 5'-CCCTT sequence in one strand. The energy of the bond breakage is preserved in the phosphotyrosine linkage between the tyrosyl residue (Tyr274) of the active site of the topoisomerase I enzyme and the 3' phosphate of the cleaved strand, making it available to be attacked by the 5'hydroxyl of the PCR fragment. The reaction is reversible and without the need of additional input of energy, as is the case for DNA ligase, since the spontaneous reformation of the phosphodiester bond regenerates both the enzyme and the DNA helix.

This is a one step cloning procedure for the direct insertion of PCR fragment into the plasmid vector pCR® 2.1-TOPO, (Invitrogen). The cloning reaction was carried out by adding 2µl of PCR product into1µl of Salt Solution (1.2M NaCl; 0.06M MgCl) and 1µl of pCR® 2.1-TOPO vector (10ng/µl plasmid DNA) making final volume to 5µl with water. The reaction was carried out by incubation at room temperature for 5 minutes.

2.5.2 Transformation of recombinant plasmids into competent E. coli cells

The TOPO vector was ligated into the competent *E. coli* cells (Invitrogen) by One Shot™ Transformation reaction. The transformed cells were grown on LB agar medium containing 50µg/ml ampicillin, 2µl X-Gal/IPTG (2.6.2)

2.5.3 One Shot™ Transformation reaction

Before transformation one vial of competent cells (Invitrogen) was thawed on ice. A vial of SOC medium was also thawed out at room temperature.

2 µl of recombinant plasmid (TOPO cloning reaction mix) was added to a vial of One Shot competent cells and gently mixed which was then incubated on ice for 30 minutes before being heat shocked at 42°C for 30 seconds in a water bath. The tube was immediately transferred to ice for 2 minutes. 250µl of room temperature SOC medium (2.6.3) was added to the vial of the cells and the cells were recovered in a shaking incubator at 37 °C and 80 rpm for 1 hour. 50µl of the transformed cells were spread onto each of LB amp/agar/X-Gal/IPTG (2.6.1 and 2.6.2) plates and the plates were kept upside- down in an incubator at 37 °C overnight.

2.5.4 Colony Purification of transformants

The colony purification was carried out by streaking the white or light blue colonies onto LB ampicillin/agar plates to grow single colonies. The plates were incubated inverted overnight at 37°C.This process is also called blue/white screening.

2.5.5 PCR screening of transformants

Colony purified colonies were directly screened by PCR. A standard 50µl PCR as described in step 2.4 was carried out with slight changes in the procedure. The reaction mixture was inoculated with the selected colony directly by using a sterile wire loop which served as a template in the reaction. A cell lysis step of 5 minutes at 94°C was added and M13F and M13R primers were used, which anneal to the flanking sites in the vector each 100bp away from the 5´ and 3´ ends of the PCR insertion. Sizes of inserts were revealed by agarose gel electrophoresis (2.3.1) and purified as described in 2.2.7.

2.5.6 Sequencing of the DNA fragments

PCR products and plasmid was sequenced using Applied Biosystem 30000 (16 capillary Genetic Analyzer) at DNA sequencing services in Advanced Biotechnology Center (ABC) Imperial College London. Both strands were sequences using both forward and reverse primers, M13F or M13R.

2.6 Bacterial Media

2.6.1 Luria Broth (LB) Medium

LB (Luria Broth) medium was generally used for the growth of bacterial colonies throughout the course of this study 10g Tryptone, 5g Yeast extract; 10g NaCl and 15g Agar was dissolved in

950ml of water. The pH of the solution was adjusted to 7.0 with NaOH and the volume was made up to 1L. The medium was then autoclaved in 500ml aliquots and stored at room temperature.

LB plates were prepared by melting 500ml of LB agar medium, cooling it down to 55°C before adding 500µl (50mg/ml) ampicillin. 25ml of this was then poured into 15cm Petri plates.

2.6.2 X-Gal and IPTG

A mixture of IPTG and X-Gal was used for the selection of the transformed colonies .A

200µl of X- Gal/IPTG was spread on the LB plates before spreading the transformed colonies on the plates. The bacterial colonies were grown at 37°C by putting the plates in inverted position in an incubator. 200 mg of X-Gal was dissolved in 9ml of dimethyl formamid (DMF), and 1ml of 20% IPTG. 200 µl of this solution was spread on each of the LB agar/ampicillin plates and allowed to dry.

2.6.3 SOC Medium

The SOC medium was made by dissolving 20g Tryptone, 5g yeast extract, 0.5g NaCl, and 15ml KCl (250mM) in 950ml of distilled water. The pH of the medium was adjusted at 7.0 with NaOH before making the volume up to 1L with distilled water

2.6.4 Liquid cultures and freezer permanents of bacterial colonies

The medium used for the growth of liquid bacterial cultures was LB liquid medium with 50 µg/ml ampicillin. The colonies were grown overnight at 37°C with shaking at 80 rpm. The liquid cultures were used to make freezer permanents. 15% (150µl/1ml of liquid cultures) dimethyl sulfoxide (DMSO) was used to make freezer permanents. The cultures were stored at -80°C. The cultures were regrown by scraping surface with a sterile stick followed by inoculation on a LB amp plate or directly into liquid culture media.

2.6.5 Tissue culture media

To induce callus from different explants of agave basic MS medium with 3% Sucrose and 8% agar was used. The following combinations were made with different growth hormones. All stock solutions for the hormones were 1mg/ml and for 100 ml of medium following amounts of hormones were added in the medium; pH was adjusted at 5.8 for all of these combinations. In first five combinations the amount of 2, 4-D (2, 4-dichlorophenoxyacetic acid) was gradually increased while keeping the amount of NAA (1-naphthylacetic acid) Constant and no BAP (6-benzylaminopurine) was added.

	2,4-D(µM)	NAA(µM)	BAP(µM)
1	2.262	2.685	0
2	4.439	2.685	0
3	9.0481	2.685	0
4	11.310	2.685	0
5	22.620	2.685	0

In second set of five combinations the amount of 2, 4-D was gradually increased and the amount of NAA was constant with an addition of a constant amount of BAP.

	2,4-D(µM)	NAA(µM)	BAP(µM)
6	2.262	2.685	4.439
7	4.4393	2.685	4.439
8	9.0481	2.685	4.439
9	11.310	2.685	4.439
10	22.620	2.685	4.439

In the third set of combinations again the same amount of 2, 4-D was used with gradual increased in every combination, the amount of NAA was doubled but remained constant for all five combinations and the BAP was eliminated from the medium

	2,4-D(µM)	NAA(µM)	BAP(µM)
11	2.262	5.73	0
12	4.4393	5.73	0
13	9.0481	5.73	0
14	11.310	5.73	0
15	22.620	5.73	0

In the fourth set of combinations the amount of 2, 4 –D and NAA remained the same as previous set of combinations but the amount of BAP was reduced to half.

	2,4-D(μM)	NAA(μM)	BAP(μM)
16	2.262	5.73	2.219
17	4.4393	5.73	2.219
18	9.0481	5.73	2.219
19	11.310	5.73	2.219
20	22.620	5.73	2.219

Four new combinations were made for the maintenance of the callus. In these combinations 5% sucrose was added with 8% agar and pH 5.85

	2,4-D(μM)	NAA(μM)	Ads(Adenine sulphate)(μM)	BAP(μM)
1	22.620	5.73	8.1446	2.219
2	22.620	5.73	0	2.219
3	22.620	5.73	0	4.4393
4	22.620	5.73	8.1446	0

The callus was growing too fast on third combination with 2, 4-D; BAP and NAA and it was hard too. The medium was modified by reducing the amount of BAP from 4.439 μM to 0.88786 μM (887.86 n moles) .This medium turned out to be the best medium for the maintenance of the callus.

2.7 DNA Sequence Analysis

Nucleotide sequences were translated into deduced peptide sequences and compared to other known retrotransposon RT sequences using online blast x, blastp and blastn algorithms of NCBI website. Multiple alignments were carried out using the ClustalW (Chenna et al., 2003) and ClustalX (Thompson et al., 1997). Phylogenetic trees were constructed and evolutionary distances calculated using the MEGA4 programme (Tamura et al., 2007). The trees and evolutionary histories were inferred using the Neighbour-Joining Method (Saitou and Nei, 1987).

Bootstrap tests of phylogeny were calculated based on 1000 replicate trees (Felsenstein, 1985) and the confidence values were shown next to the branches. The evolutionary distances for the protein sequences were computed using the Poisson correction model (Zuckerkandl and Pauling, 1965).

2.8 Amplification Fragment Length Polymorphism (AFLP)

AFLP is a DNA finger printing technique based on selective PCR amplification of restriction fragments from a total digest of genomic DNA. This technique involves restriction of DNA and ligation of oligonucleotide adapters, selective amplification of sets of restriction fragments, and gel analyses of the amplified fragments. In this study AFLP analyses was carried out as Vos et al.1995

2.8.1 Annealing of single stranded oligonucleotides to make double stranded adapters.

Stocks of *MseI* and *EcoRI* adapters were prepared to be used in AFLP and SSAP reactions.

MseI adapter stock:

MseI oligoA*	25µl
MseI oligoB**	25µl
Sterile distilled water	50µl

MseI OligoA* = 5′- GACGATGAGTCCTGAG

MseI OligoB** = TACTCAGGACTCAT- 5′

EcoRI adapter stock:

EcoRI oligoA*	25µl
EcoRI oligoB**	25µl
Sterile distilled water	50µl

EcoRI oligoA* = 5′- CTCGTAGACTGCGTACC

EcoRI oligoB** = CATCTGACGCATGGTTAA -5′

The mixture of the oligonucleotides water was incubated at 65°C for 10 minutes and placed on ice.1 µl of 1M MgOAc was added before incubating the reaction at 37v for 10 minutes followed by incubation at 25°C for 10 minutes. The adapter stocks were stored at -20°C.

2.8.2 Digestion of genomic DNA and ligation of adapters to the DNA

The digestion and ligation was carried out using *MseI* and *EcoRI* enzymes and adapters at the same time. The following double digestion and ligation reaction was carried out.

Component	Amount
Sterile distilled water	34µl
10x Ligase Buffer	5µl
Genomic DNA (1µg/µl)	5µl
EcoRI (5U/µl)	1µl
MseI (5U/µl)	1µl
EcoRI adapter (0.25µg/µl)	1µl
MseI adapter (0.5µg/µl)	1µl
T4 DNA ligase (NEB) (400U/µl)	1µl
Total	50 µl

The reaction was incubated at 37°C overnight and 5 µl of this digestion and ligation reaction was run on a 1% agarose gel.

2.8.3 Pre – amplification of restriction fragments

A pre- amplification of restriction fragments was carried out using *Mse + CG* and *Eco+G* (2.8.3) as primers. The PCR was carried out as described in step 2.4.This pre-amplification reaction was used as a template for the second PCR for the final amplification of the restriction fragments.

2.8.4 Preparation of radio-labeled primers

Radio-labeled primers were used for the final amplification of restriction fragments in AFLP and SSAP. In AFLP *EcoRI* Primer was labeled while in SSAP the retrotransposon specific primer was labeled with γ^{33} P ATP. The following labeling reaction was carried out for 10 PCR reactions.

Sterile distilled water	6.41 µl
10x T4 kinase buffer	1.0 µl
T4 polynucleotide kinase (10u/ µl)	0.25 µl
γ^{33} P ATP	1.0 µl
Primer *(0.05µg/ µl)	1.34µl

Primer *= Eco+G in case of AFLP and retrotransposon specific primer in case of SSAP

The reaction was incubated at 37°C for 3 hours and at 70°C for 10 minutes and stored on ice before use.1 µl of this labeling reaction was used in the second PCR of AFLP or SSAP depending up on the PCR reaction. A stock enough for all of the samples was usually made.

2.8.5 Amplification of the restriction fragments using radio-labeled primer

The pre-amplified restriction fragments were amplified and labeled using *Eco+G* radio- labeled primer in combination with *Mse+CG*. A touchdown PCR was carried out with starting annealing temperature of 65°C reducing 0.7°C each cycle for first 12 cycles and remaining constant for the rest of the reaction. The total number of cycles was 35.The rest of the conditions was same as a standard PCR described in 2.4. The PCR products were resolved on a 6% polyacrylamide urea gel as described in 2.3.3.

2.9 Sequence Specific amplification Polymorphism (SSAP)

Sequence specific amplification polymorphism is a retrotransposon based molecular marker system. The methodology of SSAP is similar to that of AFLP as DNA is digested with two restriction enzymes and adapters are ligated to the restriction fragments. These fragments are pre-amplified using adapter specific primers. The last step of SSAP differs from AFLP as it amplifies the DNA fragments between a LTR and the nearby restriction site using a radio labeled LTR specific primer (170 A1/3 5´-GATTGTAACCTTGGGCC-3´) and an adapter specific primer (Eco or Mse etc). The SSAP was carried out as Syed et al.2006 and Bousios et al.2007.

The ligation of primers, digestion of DNA and the pre-amplification was carried out as described above in 2.8. Retrotransposon specific primer was labeled with $^{\gamma 33}$P ATP (2.8.4). The rest of the reaction was same as described in 2.8. A touch down PCR was carried out and the PCR products were resolved on a polyacrylamide gel as in 2.3.3. The primers used in this experiment were taken from Bousious et al.2007.

2.10 Inter Retrotransposon Amplification Polymorphism (IRAP)

Inter retrotransposon amplification polymorphism is retrotransposon based marker system which uses retrotransposon LTR specific primer to amplify fragments between two adjacent LTR

retrotransposons. The products of IRAP can be visualized directly by agarose gel electrophoresis (2.3.3). However if 33p labeled primers (2.8.4) are used the products will be resolved by a polyaccrylamide gel electrophoresis and visualized by autoradiograph. The PCR products were separated on a polyacrylamide gel and visualized by an auto radiograph.

In this study simple PCR reaction with non radio labeled primer was carried out using genomic DNA and retrotransposon specific primer to make a pre IRAP and this reaction was used as a template to do another PCR using the same primer again. The fragments amplified were resolved on a 2% high resolution agarose as described in 2.3.2.

Oligo number	Name of primer	Primer sequence
408	A1A IRAP	5´-CAATGAATATGTGGTACT-3´
409	A1B IRAP	5´-GATTGTAACCTTGGGCCCAACA-3´
453	A17 IRAP	5´-GCTTAAAGGGCCAACAG-3´

2.11 Slot Blotting and Hybridization

Slot blotting is a technique is a simplified form of Southern or northern blotting as it does not involve gel electrophoresis. In this technique the biomolecules are directly transferred to a membrane by vacuum using a slot blotter and gives a measure of heterogeneity levels. The probes are hybridized to the membrane by conventional southern hybridization. In the present study the slot blotting was used for the estimation of retrotransposon copy number and heterogeneity according to the procedure described in the following steps.

2.11.1 Preparation of probes

The DNAs used as probes were quantified using a digital Biophotometer 6131 by Ppendr; Hamburg Germany at a 260.100ng of DNA was boiled for 5minutes and cooled immediately on ice. The probe was labeled using Amersham AlkPhos Direct^{TM} Labeling and Detection system RPN3690 (GE healthcare UK Ltd). The labeled probe was stored at -20°C in 50% glycerol.

2.11.2 Preparation of slots blots

The slot blotting was carried out by using a vacuum slot blotter. A piece of the membrane Hybond-N+ (GE Healthcare, UK Ltd. RPN 203B) was cut according to size f the blotter. The

membrane was dipped in 2× SSC which was made by diluting 20×SSC (175.3g NaCl, 88.2g sodium citrate, water up to a total volume of 1L and pH of 7.0). The apparatus was attached to the vacuum pump and the vacuum was turned on a few minutes before loading the samples. The DNA to be loaded was denatured in equal volume of denaturation solution (0.2 MNaOH 2MNaCl) for 5 minutes before loading to the slots. In another row of slots we added a range of concentrations of control DNA. The amount that is added depends upon the size of the fragment and the genome size of the species in pg. The solution was transferred at the membrane b vacuum. The membrane was washed by adding 500µl of 2 x SSC to the slots. The blots was then taken out of the blotter and DNA was fixed to the membrane by ultra violet light by placing the membrane on a UV tans-illuminator (DNA side down) at a wavelength of 302 nm for 1minutes.

2.11.3 Hybridization

Hybridization was carried out hybridization buffer provided with Amersham AlkPhos Direct™ Labeling and Detection system RPN39 (GE healthcare UK Ltd.). The working hybirdsation buffer was made by adding 0.5 M NaCl and 4% (W/V) blocking regent also provided with the kit. The blot was pre hybridized by incubating in the hybridization over for 1 hour at 55°C with pre warmed hybridization buffer (55°C). The amount of buffer required for 1cm^2 of membrane is 0.125ml (10ml for 80cm^2). After the pre hybridization step the buffer was poured out of the hybridization bottle in a beaker and labeled probe was added and mixed. The mixture was poured back into the hybe-bottle and left for incubation in the rotating hybe-oven at 55°C for 12-15 hours.

2.11.4 Washes

After the hybridization for 12-15 ours the blot was washed twice in 100ml of pre warmed (55°C) primary wash buffer (60g Urea 5ml of 10% SDS, 50ml of 5M sodium phosphate pH7, 500µl of 1M $MgCl_2$, 1g blocking reagent, water up to total volume of 500ml) for 10 minutes at 55°C with shaking. The blot was then washed twice in 1x secondary wash buffer (25ml of 20x secondary wash buffer, 1ml 1 M $MgCl_2$ total volume to 500ml with water) for 5 minutes. The blot was subjected to a range of stringency washes for 1 hour at 55°C, 60°C and 65°C in 1x secondary wash buffer; and at 65°C in 0.1x secondary wash buffer for 1 hour.

** 20x secondary wash buffer (60g Trizma base Sigma T1503 and 56g NaCl, 450 ml water, adjusting pH to 7 and making the volume up to 500ml).

2.11.5 Signal generation and detection by CDP star detection

Detection reagent CDP star was provided with the Amersham ALkPhos Direct^{TM} Labeling and Detection system. The excess wash buffer was drained from the blot and it was placed in a plastic bag with DNA side upwards. Detection reagent supplied with kit was poured on the blot by using pipette (2.4ml/80cm^2), excess detection reagent was drained off and the bag was sealed I a sealer. The blot was placed in an autoradiograph cassette and piece of film was placed on top of the blot. The blot was exposed for an hour at room temperature and the film was developed in a developer *Konica SRX 101 A.*

2.11.6 Northern Blotting and Hybridization

The northern blotting is a technique used in molecular biology for the expression studies. In this technique RNA is used instead of DNA as well as RNA probe could be used for the hybridization.6µl of RNA was loaded with an equal volume of loading buffer on a formaldehyde gel and the gel was run for 1 and a half hours at 70V. On the other hand the blotting apparatus was prepared. A piece of Hybond –N+ membrane (GE Healthcare, UK Ltd, RPN 203B) was dipped in 20x SSC buffer (17.3g NaCl, 88.2g sodium citrate, water up to a total volume of 1L and pH of 7.0). The gel was placed on the apparatus wit a piece of 3mm Whatman paper underneath going into the lower tank of the apparatus. The lower tank of the apparatus was filled with 20x SSC buffer. A piece of saran wrap was cut in the middle according to the size of the gel and placed on Whatman paper

CHAPTER 3

Isolation and characterization of retrotransposon reverse transcriptase sequences from *Agave tequilana.*

3.1: Introduction

Retrotransposons are the most abundant class of eukaryotic transposable elements consisting of LTR retrotransposons and non- LTR retrotransposons (Bennetzen, 1996; (Grandbastien, 1992; Kumar and Bennetzen, 1999; Kunze et al., 1997). The LTR retrotransposons can be further subclassified into two groups, the *Ty1-copia* group and the *Ty3-Gypsy* group on the basis of their degree of sequence similarity as well as the order of their genes (Kumar and Bennetzen, 1999; Wicker and Keller, 2007). The two LTR retrotransposon groups have also been named as Pseudoviridae (*Ty1-copia*) and Metaviridae (*Ty3-gypsy*) (Eickbush and Jamburuthugoda, 2008).

The internal genomic structure of LTR retrotransposons is similar to that of retroviruses except for the absence of *env* gene in most elements. All LTR retrotransposons contain gag and *pol* genes. The *gag* gene is the most variable but typically encodes major structural and nucleic acid binding domains which may be involved in reverse transcription. The *pol* gene encodes the various enzymatic domains like proteinase (PR), reverse transcriptase (RT), RNaseH domain and integrase (IN) domain. (Boeke and Corces, 1989; Eickbush and Jamburuthugoda, 2008; Sandmeyer et al., 1990).In the *Ty1-copia* retrotransposons IN domain is located amino terminal to the RT and RNaseH domains while in *Ty3-* gypsy elements the IN domain is located as in retroviruses at the carboxyl- terminal end of the *pol* gene (Eickbush and Jamburuthugoda, 2008).
Ty1-*copia* group retrotransposons are the major group of elements in higher plants, varying greatly in copy number over relatively short evolutionary timescale. Therefore, they are one of the most important factors affecting the structural evolution of the plant genomes (Gribbon et al., 1999) and contribute to the genetic diversity of their host genomes (Hernandez et al., 2001; Linares et al., 2001; Price et al., 2002; Verries et al., 2000). Ty1-*copia* retrotransposons possess

several important characteristics that make them suitable for studying the structure and organization of plant genomes. They have been widely characterized for several plant species showing broad insertional patterns, heterogeneity, and sequence variability (Brandes et al., 1997; Ellis et al., 1998; Flavell et al., 1992b; Garber et al., 1999; Hirochika et al., 1992; Pearce et al., 1997; Pearce et al., 1999; Pearce et al., 1996c; Waugh et al., 1997). These retrotransposons have been characterized in wide variety of plants including monocots and dicots for example *BARE1* in barley (Manninen and Schulman, 1993), Tnt1 in tobacco (Grandbastien et al., 1989), Tto1-3 in tobacco (Hirochika, 1993), RIRE1 (Sasaki et al., 2002) and Tos17 in rice (Hirochika et al., 1996b) Tal3 in *Arabidopsis* (Konieczny et al., 1991), *Opie-1* in maize (SanMiguel et al., 1996), *Athila* in *Arabidopsis* (Pelissier et al., 1995) and *Retrolycl* in tomato (Costa et al., 1999).

3.1.1: Use of reverse transcriptase (RT) to study retrotransposons.

For the last twenty years different attempts have been made to use RT sequences to determine the phylogenetic relationship of retrotransposons (Eickbush and Jamburuthugoda, 2008). In plants the RT domain has been used to detect retrotransposons and to assess their distribution and evolution. The detailed characterization of different plant taxa with respect to the content, variability and physical distribution of retrotransposons made a major contribution to the understanding of host genome organization and evolution (Friesen et al., 2001). The presence of conserved amino acids in the reverse transcriptase domain has enabled degenerate oligonucleotide primers to be designed that have widely been used to amplify reverse transcriptase sequences of retrotransposons in diverse species including single celled algae, bryophytes, gymnosperms and angiosperms (Flavell et al., 1992b; Grandbastien, 1998; Kumar et al., 1997; Pearce et al., 1996b; Voytas and Ausubel, 1988; Xiao et al., 2004). Due to its conservation across species and its role in the retrotransposon research, RT was the best domain to be used to investigate retrotransposons in blue agave.

To isolate and characterize Ty1-*copia* retrotransposons in *Agave tequilana,* degenerate primers specific to Ty1-*copia* retrotransposon reverse transcriptase were used to amplify RT fragments by polymerase chain reaction (PCR).The sequence of upstream primer is 5´-ARCATRTCRTCNACRTA-3´ and the sequence of down stream primer is

5´-AARGCNNGNYTNGTNGCNMARG -3´.The down stream primer corresponds to the motif YVDDM which is a highly conserved motif of retrotransposons reverse transcriptase while

upstream primer corresponds to the KARLVA(QK)G motif of retrotransposon reverse transcriptase and is coserved in *Ty1-copia* retrotransposons but it is not conserved in all of the retrotransposon so it only helps to isolate *Ty1-copia* retrotransposons (Flavell et al., 1992a). These two primers were used to isolate Ty1-copia retrotranspsons by Flavell et al (1992) and have been extensively used for the isolation of retrotransposons in a wide variety of plants. To evaluate the validity of these primers retrotransposon reverse transcriptase sequences from different plants with special emphasis to rice and *Arabidopsis* were obtained from the sequence data bases using NCBI blast search.

Figureure A : Clustal w alignment of reverse transcriptase sequences from different plants. Reverse transcriptase sequences from different plants were aligned using clustal w alignment.The KARLVAKG motif is highlighted as red while YVDDML motif is highlighted as green.

Figure A shows the clustal w alignment of the sequences from rice ,arabidopses, tomato, tobacco and model legume *Medicago truncatula* clearly showing the conserved KARLVAK(Q)G and YVDDML motifs of reverse transcriptase domain.

The amplified fragments were subcloned into plasmid vector pCR® 2.1-TOPO, (Invitrogen) so that the isolated fragments could be sequenced. The RT fragments were amplified by PCR using *TaqPolymerase* which has a nontemplate dependent terminal transferase activity that adds a single deoxyadenosine (A) to the 3´ end of the PCR products while the vector used in this work has asingle overhanging deoxithymidine (T) which allows PCR insert to ligate to the vactor (Shulman 1991,1994). So the RT fragments amplified by PCR using degenerate primers corresponding to the conserved motifs of reverse transcriptase domain were inserted into the pCR® 2.1-TOPO vector. This vector contains priming sites for M13 reverse (5´-

GTCATAGCTGTTTCCTG-3) and M13 forward (5´-GTAAAACGACGGCCAG-3´) primers 100bp up and downstream of the PCR product insertion site. So when these primers are used a fragment containing the PCR product plus 100bp on both sides of the product is amplified. The isolated fragments were subcloned and sequenced as described in section 2.5 of chapter 2.

The main aim of this chapter was to isolate and characterize Ty1-copia retrotransposonss from *Agave tequilana* and study their phylogenetic relationship other retrotransposons.

The isolation and characterization of Ty1-copia retrotransposons revealed that they occure in the form of subgroups of closely related heterogeneous elements in *Agave tequilana*.

3.2: Results

This chapter presents the isolation and characterization of *Ty1- copia* retrotranspososn reverse transcriptase sequences in *Agave tequilana*. The alignments of deduced polypeptide RT sequences revealed four major subgroups of closely related heterogeneous sequences as well as individual ungrouped sequences. Phylogenetic analyses showed that different subgroups of sequences cluster together pointing towards a possible amplification of these retrotransposons in the recent past. In this piece of work approximately 70 clones were selected for sequencing and 42 of these sequences were included in the analysis.

3.2.1: Characterisation of Ty1-*copia* retrotransposon RT (reverse transcriptase) sequences.

The characterization could start once the sequence data was available, targeting at the conserved motifs KRLVAK(Q)G and YVDDM which flank the internal RT domain. The nucleotide sequences were translated and the deduced peptide sequences were compared to the known retrotransposon reverse transcriptase sequences. The available peptide sequences were manually compared to the other retrotransposons like *SIRE-1, Tto1, Tnt1, OPIE1 , BARE-1, COPIA, PDR1 and Ta1* and conserved motifs shared by all of them like YMEQP and SLYGLK were located. The identified RT sequences were used to perform a blast search on NCBI- BLAST for their comparison to the other known retrotransposon reverse transcriptase sequences. The blast searches of the newly isolated retrotransposon reverse transcriptase (RT) sequences showed great similarity to the other retrotransposons. The sequences were manually checked for stop codons and frameshifts. Some of the sequences had stop codons or frameshifts or both of them together.

Figure 3.1 shows a ClustalX (Thompson et al.1997) multiple sequence alignment of RT peptide sequences, structure of a standard LTR retrotransposon with the position of RT gene and four major subgroups. The multiple sequence alignment revealed a heterogeneous population of RT sequences which are closely related to each other (Figure 3.1). The heterogeneous population of retrotransposon RT sequences consisted of four major clusters of sequences namely Teq1, Teq2, and Teq3 and Teq24 subgroups as well as ungrouped individual sequences like Teq6, Teq22, Teq29, Teq31 and Teq41.Apart from four major subgroups there was also a small subgroup of three sequences Teq8, Teq20 and Teq21. The major subgroups are named after the sequence with lowest number (1, 2, 3) in the subgroup like Teq1or Teq24 (Figure 3.1)

The population of Ty1-copia RT sequences is evenly divided among four subgroups with Teq2 being the largest subgroup containing 10 (23.8%) sequences followed by Teq1 and Teq3 with 9(21.4%) and 8 (19%) sequences respectively, while Teq24 contains 7 (16%) sequences. In other words around 80.95% sequences are shared by four major subgroups. These subgroups show a peptide sequence identity of 94.35% (Teq1), 87.88% (Teq2), 73.5% (Teq3), and 93.38% (Teq24) respectively, showing high level of conservation within the clusters of sequences thus making them distinct subgroups of closely related retrotransposon RT fragments. Apart from the major groups there is a small group of three sequences namely Teq8, Teq20 and Teq21 with 100% peptide sequences identity and 97.6% nucleotide sequence identity among themselves.

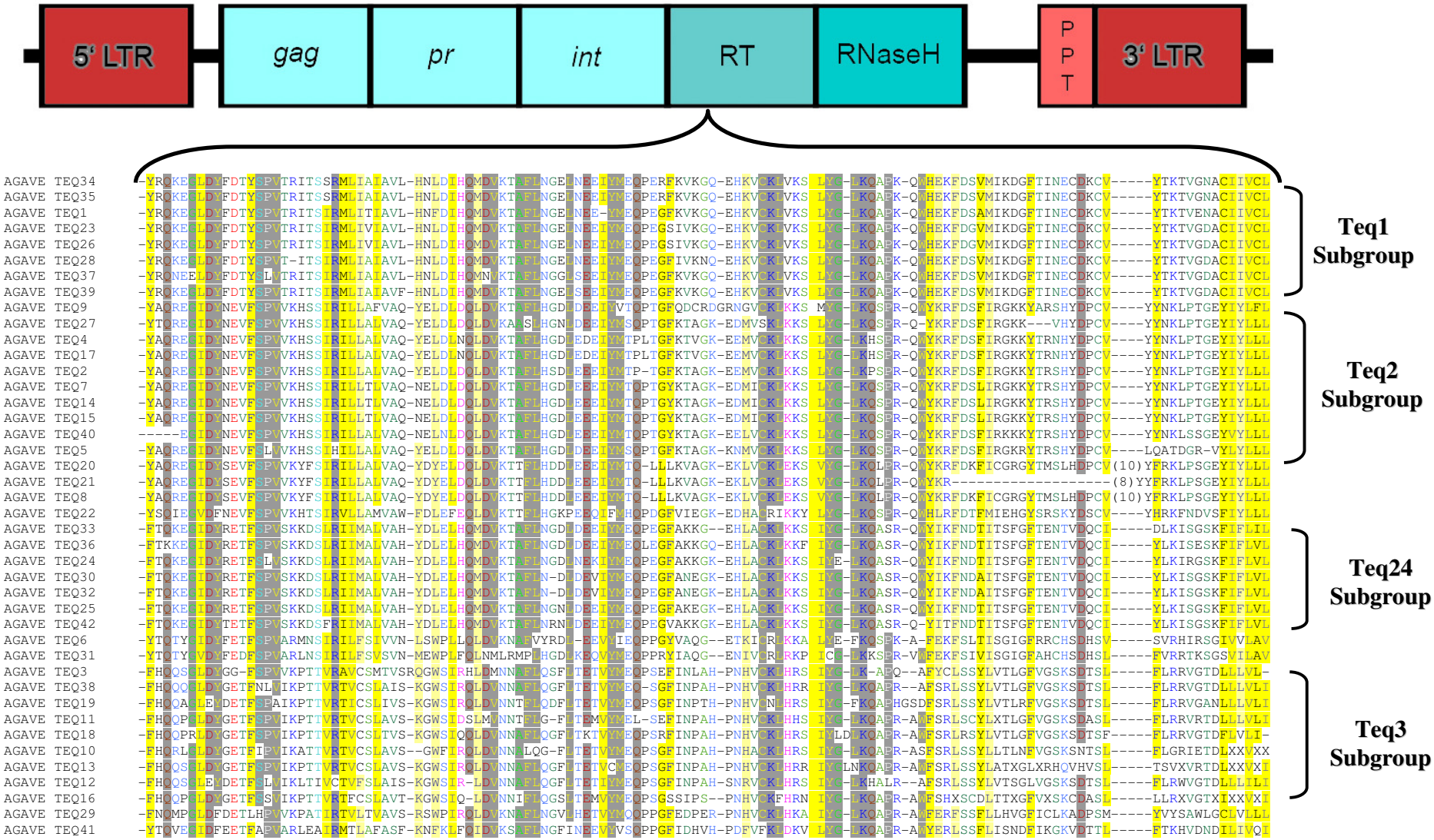


Figure 3.1: Clustal X multiple sequence alignment of Ty1-copia retrotransposon reverse transcriptase (RT) peptide sequences. Diagrammatic illustration of a typical Ty1-copia retrotransposons highlighting different parts of its coding domain and LTRs are shown on the top of the alignment. Highly conserved amino acids are highlighted by the colors and major subgroups of closely related peptide sequences are illustrated by the brackets on the right of the Figureure. Ungrouped individual sequences are not enclosed in the brackets.

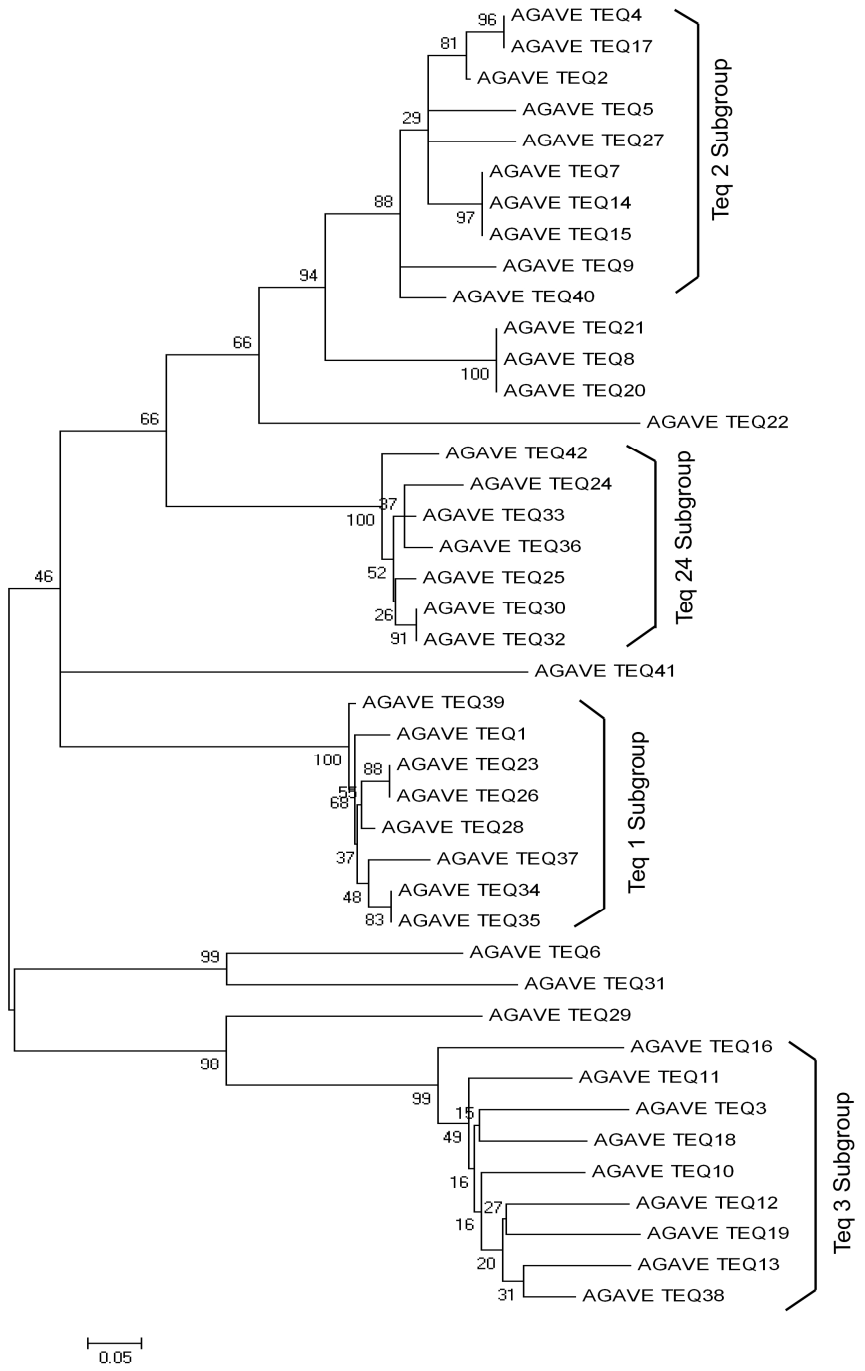


Figure: 3.2A Phylogenetic relationships of Ty1-copia retrotransposon reverse ranscriptase (RT) sequences isolated from *Agave tequilana*

The evolutionary history was inferred using the Neighbor-Joining method. The bootstrap consensus tree inferred from 1000 replicates is taken to represent the evolutionary history of the sequences analyzed. Branches corresponding to partitions reproduced in less than 50% bootstrap replicates are collapsed. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (1000 replicates) are shown next to the branches. The evolutionary distances were computed using the Poisson correction method and are in the units of the number of amino acid substitutions per site. All positions containing gaps and missing data were eliminated from the dataset (Complete deletion option). There were a total of 80 positions in the final dataset. Phylogenetic analyses were conducted in MEGA4.The major subgroups are shown by the brackets.

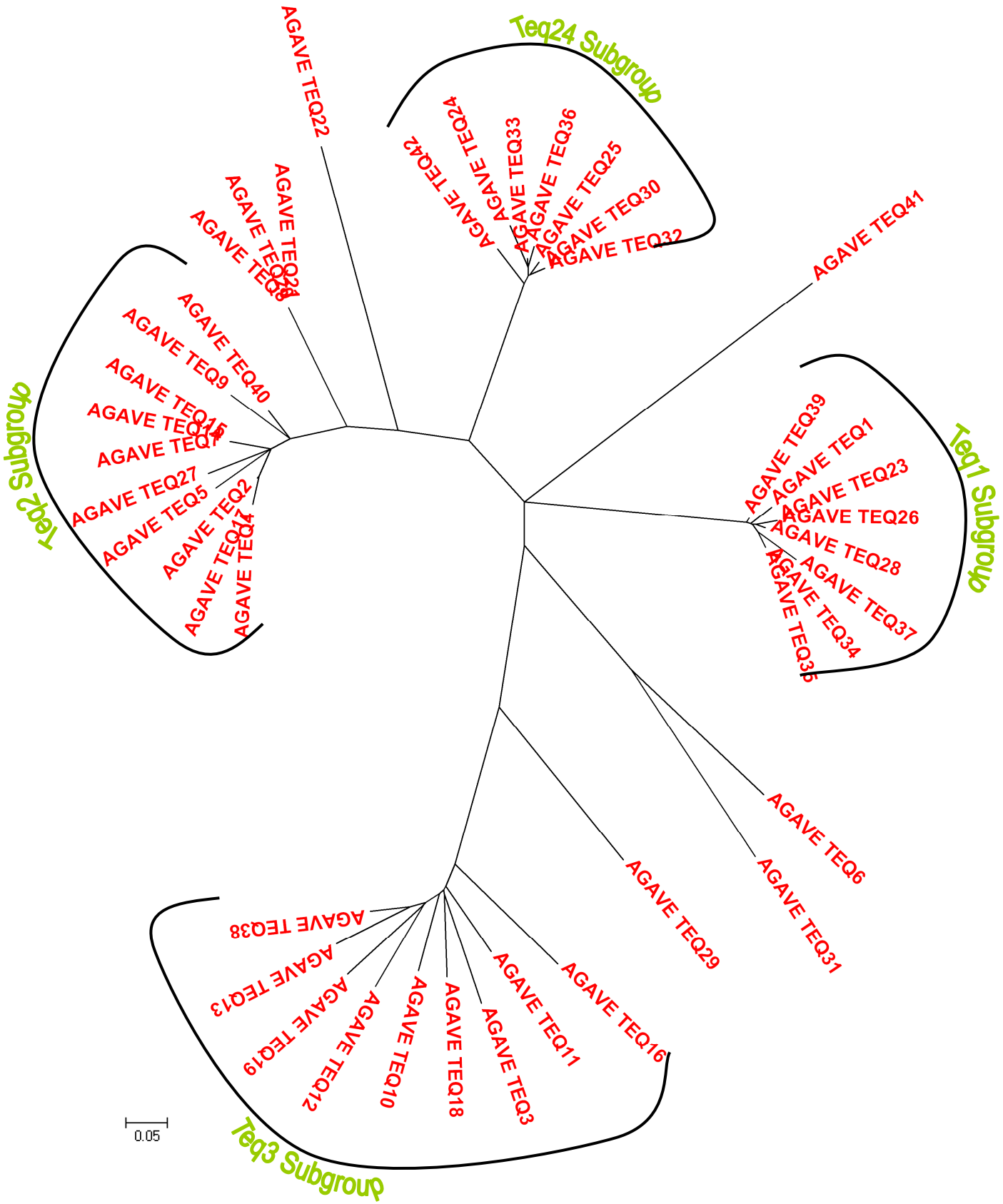


Figure: 3.2B Unrooted tree of Ty1-copia retrotransposon RT sequences from agave. Different subgroups cluster occur in the form of clusters of sequences. Four major subgroups are highlighted with black lines around them and the ungrouped sequences branch out separately with long branches.

The heterogeneity among different subgroups and individual RT sequences was seen and their evolutionary relationship was established by constructing a phylogenetic tree based on the Ty1-*copia* peptide reverse transcriptase sequences (Figure 3.2A, 3.2B). A bootstrap test was performed with 1000 replicates to assess the reliability of the inferred phylogenetic tree. The branching pattern of different subgroups was strongly supported by the confidence level of the tree, offering a high bootstrap value for each of the subgroup (Figure 3.2A and 3.2B). The bootstrap confidence value was 100% for Teq1 and Teq24 subgroups and 99% and 94% for Teq2 and Teq3 subgroups respectively, while it was 100% for Teq8 subgroup. The ungrouped elements exhibited low confidence values as well as relatively longer branch lengths thus separating themselves from the rest of the elements. Teq22 and Teq41 showed longest branch lengths as compared to the other ungrouped individual sequences indicating towards higher divergence. Additionally phylogenetic analysis of *Ty1-copia* nucleotide sequences was also carried out and a neighbour joining tree was constructed and tested with a bootstrap value of 1000 replicates as shown in Figure 3.3. To test the reliability of the trees different tests were carried and no significant changes were found in branching pattern and the bootstrap confidence value of the trees with other methods. However Teq38 which grouped together with Teq3 elements in the peptide alignment and neighbour joining peptide tree lies itself with Teq1 elements instead in the tree based on nucleotide sequences Figure 3.3

3.2.2 Sequence analysis and evolutionary relationship of Teq1 subgroup elements

Teq1 subgroup is one of the four major subgroups of Ty1-*copia* retrotransposon RT sequences. This subgroup consists of eight RT sequences all of which possess an uninterrupted open reading frame. These sequences are closely related to each other with an average 94% nucleotide and 94.35% peptide sequence identity among them.

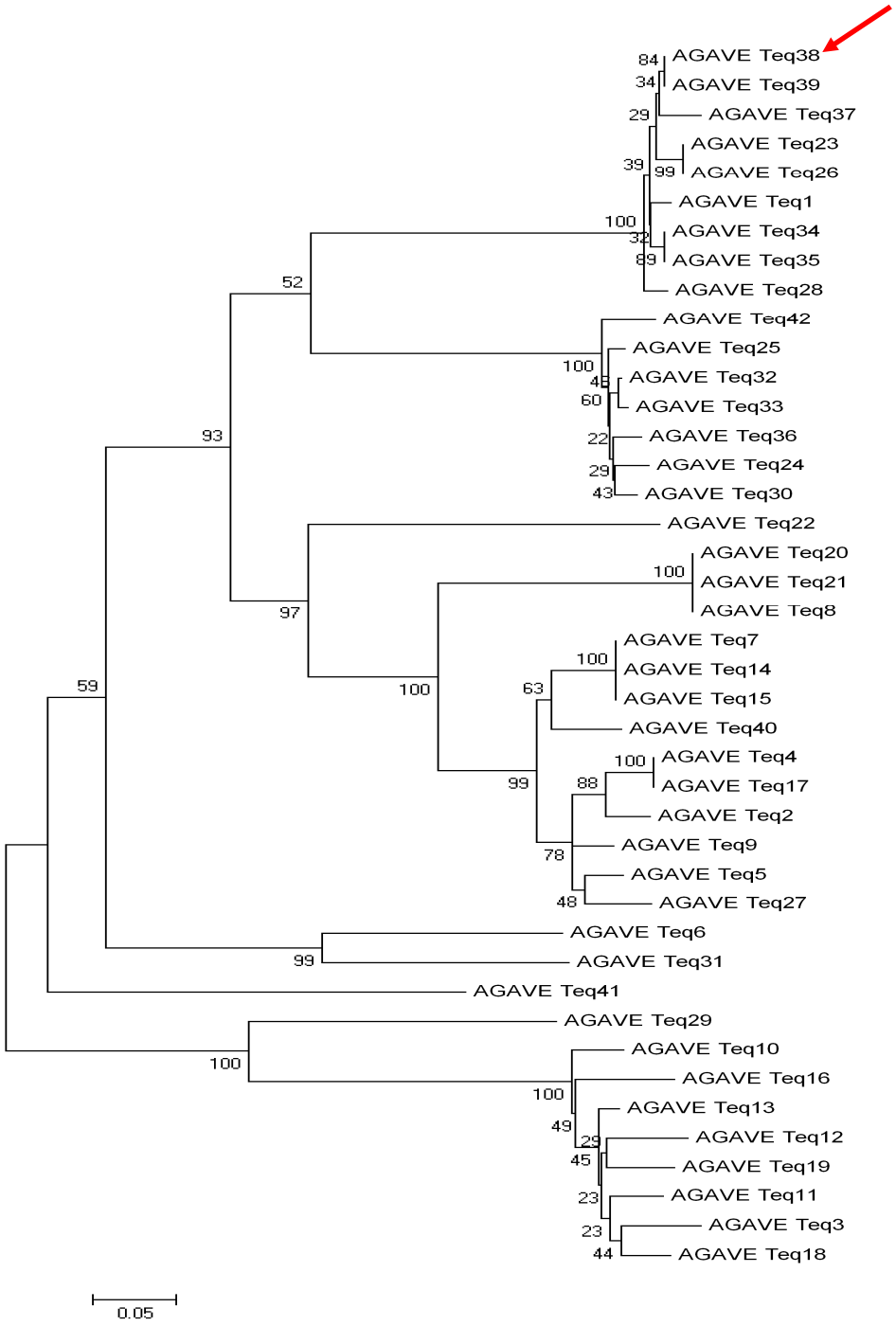


Figure: 3.3 Phylogenetic relationships of retrotransposon RT nucleotide sequences isolated from *Agave tequilana*

The evolutionary history was inferred using the Neighbor-Joining method .The bootstrap consensus tree inferred from 1000 replicates is taken to represent the evolutionary history of the nucleotide sequences analysed. Branches corresponding to partitions reproduced in less than 50% bootstrap replicates are collapsed. The percentage of replicate trees in which the associated retrotransposon RT sequences clustered together in the bootstrap test (1000 replicates) is shown next to the branches. The tree is drawn to scale, with branch lengths in the same units as those of the evolutionary distances used to infer the phylogenetic tree. The evolutionary distances were computed using the Maximum Composite Likelihood method and are in the units of the number of base substitutions per site. All positions containing gaps and missing data were eliminated from the dataset (Complete deletion option). Phylogenetic analyses were conducted in MEGA4. Red arrow indicates teq38 grouping with Teq1 subgroup elements.

The phylogenetic relationship of Teq1 subgroup elements is shown by the neighbour joining tree of Teq1 RT peptide sequences (Figure 3.2A) showing a high bootstrap confidence value of 100% that makes it a distinct group of the sequences, while Figure: 3.4) shows the relationship of Teq1 sequences separately as it is the branch of the phylogenetic tree showing Teq 1 sequences only. A closer look at the multiple alignments of the sequences revealed that all of the Teq1 subgroup elements share the conserved motifs of RT and possible heterogeneity among the elements of this subgroup is of 0- 9%. However Teq37 is slightly different from the rest of the group as it shows a relatively longer branch length in the phylogenetic tree (Figure: 3.4) suggesting an early divergence from the rest of the group.

On the basis of short branch length on the phylogenetic tree (Figure 3.2A), the presence of an open reading frame in all of the Teq1 sequences and high nucleotide as well as peptide sequence identity, Teq1 subgroup could potentially be a high copy number subgroup of Ty1-*copia* retrotransposon which may have recently been actively replicating.

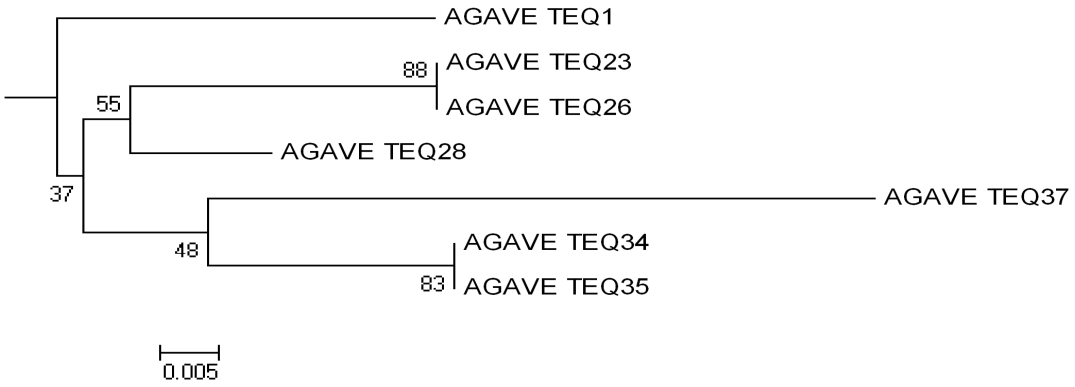


Figure: 3.4 Phylogenetic relationships of Teq1 elements
Phylogenetic relationship of Teq1 elements shown by the sub-tree .The evolutionary history was inferred by neighbor joining method and sub- tree was taken from the neighbor joining tree of all of the sequences. Red arrow highlights Teq37 with a longer branch length.

3.2.3 Sequence analysis and evolutionary relationship of Teq2 subgroup elements

Teq2 subgroup is the largest group of retrotransposon reverse transcriptase sequences. There are at least 10 members of Teq2 subgroup with high nucleotide (76.47%) as well as peptide (87.88%) sequence identity among them. In addition to the high sequence identities 70% of the members of this group also posses uninterrupted open reading frames .The average GC content of this subgroup is 44.36% (Table 3.1)

The phylogenetic relationship of Teq2 elements is presented by neighbour joining tree (Figure 3.2A) .It can be seen from the tree that Teq2 subgroup is a diverse group with differences in the brach lengths of the elements , however a high bootstrap confidence value of 99%, which makes it a distinct subgroup. Teq5 and Teq27, two members of the subgroup have a longer branch length showing an earlier divergence from the rest of the group. However some of the Teq2 elements have very short branch lengths (see Teq2, Teq4, Teq7, Teq14 and Teq15) (Figure 3.2A) Moreover about 30% of Teq2 subgroup RT sequences have stop codons suggesting that some members of this group are defective and might not be able to replicate properly. Nevertheless short branch length with high bootstrap confidence value as well as high sequence (nucleotide and peptide) similarity among Teq2 elements reverse transcriptase sequence indicates that Teq2 subgroup is a potentially active and abundant group of Ty1-*copia* retrotransposons. Short branch length particularly suggests a possible activation of these elements in recent past and the difference in the branch length indicates the heterogeneous nature of the group. A closer look at the multiple alignments of the sequences and percentage identity of different members of the subgroup shows that these elements have 0-23% heterogeneity among them. In conclusion Teq2 subgroup consists of a population of closely related but heterogeneous Ty1-*copia* retrotransposons, some of which might be active in recent past. These elements might also be present in high copy number in *Agave tequilana* genome.

3.2.4 Sequence analysis and evolutionary relationship of Teq3 subgroup elements

Teq3 subgroup consists of nine newly isolated reverse transcriptase (RT) sequences. These sequences share a high nucleotide sequence identity (87.6%) as well as peptide sequence identity (73.5%), however the peptide sequence identity is slightly lower than the Teq1, Teq2 and Teq24 subgroups. Moreover Teq3 subgroup reverse transcriptase subgroup nucleotide sequences are 43.26% GC rich on average (Table 3.1).

The phylogenetic and evolutionary relationship of Teq3 subgroup elements is described by neighbour joining tree (Figure 3.1) showing a high bootstrap confidence value (99%) making

Teq3 group a distinct cluster of closely related sequences. The branch length for most of Teq3 sequences is almost the same apart from Teq16 which has a longer branch length. As a whole Teq3 elements possess longer branch lengths as compared to the other subgroups of reverse transcriptase (RT) sequences (Figure 3.5). Interestingly the majority (77%) of Teq3 sequences contain either a stop codon or frameshift or both, suggesting that most of Teq3 elements might be defective and unable to replicate. Although most of the Teq3 subgroup element analysed here are defective, however they have a high nucleotide as well as peptide sequence similarity and a high bootstrap confidence value (94%) on neighbour joining tree (Figure 3.2). They also form a cluster of closely related sequences (Figure 3.1) as shown in the multiple alignments of these sequences. All of these characteristics suggest their abundance in the genome of agave and a possible activation in past. Nevertheless the presence of stop codons as well as frameshifts points towards a defective coding region making most of Teq3 elements incapable of replication, but this could also be the result of a recent burst of activation which might have produced some partially or fully defective Teq3 type *Ty1-copia* elements along with active elements in the genome of blue agave

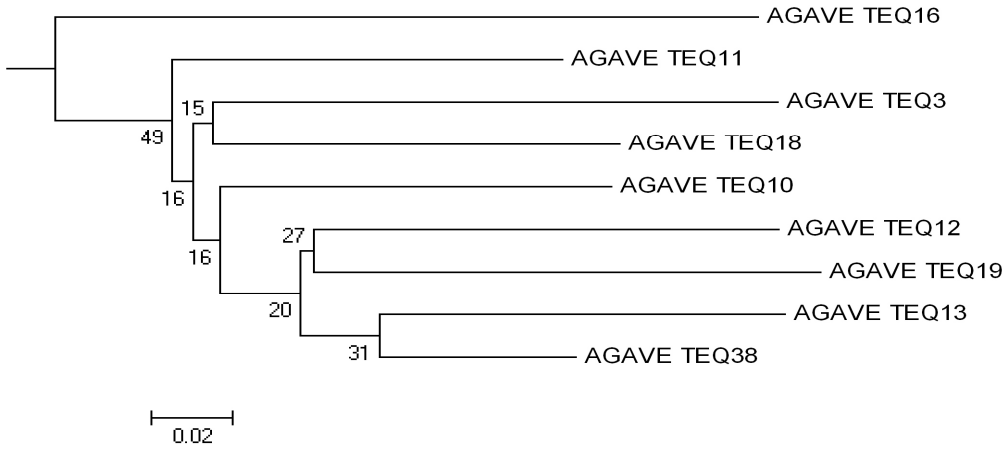


Figure 3.5 Phylogenetic relationship of Teq3 elements.
Phylogenetic relationship of Teq3 elements shown by the sub-tree .The evolutionary history was inferred by neighbor joining method and sub- tree was taken from the neighbor joining tree of all of the sequences.

3.2.5 Sequence analysis and evolutionary relationship of Teq24 subgroup elements

Teq24 subgroup is the smallest of major groups comprising of seven closely related reverse transcriptase sequences which are highly similar to each other with respect to their nucleotide sequences (95%) as well as peptide sequences (93.5). In addition to high sequence identity these sequences are 34.66% GC rich and do not contain any interruptions in the coding domain of their reverse transcriptase.

The evolutionary relationship of these sequences is shown by neighbour joining phylogenetic tree (Figure 3.2) which shows a branching pattern strongly supported by a high bootstrap confidence value (100%) which makes them a distinct subgroup of elements. All of Teq24 subgroup sequences have very short branch lengths in the phylogenetic tree suggesting a possible activation of these elements in the recent past; moreover the presence of an open reading frame in all of these sequences shows that they can possibly be autonomous elements.

A high sequence identity, short branch length with high bootstrap confidence value and presence of an open reading frame in all of Teq24 subgroup sequences indicates that these sequences might form a population of closely related and highly abundant heterogeneous *Ty1-copia* elements, in the genome of blue agave.

Subgroup	Average GC content %	Nucleotide Identity %	Peptide Identity %	Presence of ORF %
Teq1	**34.49**	**94**	**94.35**	**100**
Teq2	**44.36**	**76**	**87.88**	**70**
Teq3	**43.26**	**87.6**	**73.50**	**30**
Teq24	**34.66**	**95**	**93.38**	**100**

Table 3.2 Average GC content, nucleotide and peptide sequence identities and frequency of ORF of major subgroups shown in percentage. GC content was calculated for every sequence and a simple average was calculated for each group. Average sequence identities (nucleotide and peptide) were calculated for the major subgroups and percentage of ORF was calculated by dividing the number of sequences with ORF with number of total sequences and multiplying with 100.

3.2.6 Sequence analysis and evolutionary relationship of ungrouped individual elements

The population of newly isolated RT sequences from agave also contains some ungrouped sequences with no significant homology to the four major subgroups of elements. Among these ungrouped elements three elements are closely related to each other which can potentially form another small subgroup. Teq8, Teq20 and Teq21 show great sequence identity (100% peptide sequence identity) with each other and also contain the conserved amino acids commonly found in RT sequences but these sequences are slightly longer than the rest of the population (Figure 3.1). These sequences also cluster together on the phylogentic tree with a bootstrap confidence value of 100% (Figure 3.2A and 3.2B) Two of these three elements contain frameshifts but their high sequence identity and grouping together on the neighbour joining tree with high bootstrap confidence value suggests that they might possibly be another subgroup of elements. The maximum sequence identity of these sequences is 76% (between Teq21 and Teq2) with the members of Teq2 subgroup and they show a bootstrap value of 94% with the rest of Teq2 subgroup elements. Due to high bootstrap value on the phylogenetic tree these three elements could be considered as a part of Teq2 subgroup but low sequence identity with the Teq2 subgroup elements suggests that they may belong to a related but different subgroup of elements.

Teq22 is another individual sequence which is highly conserved with respect to its reverse transcriptase peptide sequence when compared to the other *Ty1-copia* retrotransposons (Figure 3.1). Teq22 also contains an open reading frame without any stop codons or frameshifts and possesses a long branch on the neighbour joining tree (Figure 3.2A) indicating a possible earlier divergence from the rest of the population.

The ungrouped population of reverse transcriptase sequences isolated from agave also includes Teq41 which contains an open reading frame like Teq22. In the neighbour joining phylogenetic tree (Figure 3.2A) it does not group with any of the other sequences, in addition to its separate position on the phylogenetic tree Teq41 also shows the highest phylogenetic distance value (0.4356) even higher than Teq22 which also shows a high distance value (0.354). Teq41 is also a sequence with a long branch and could potentially be a high copy number element in the genome of *Agave tequilana*. Teq6 and Teq31 are the other individual sequences which do not cluster with the rest of the sequences but they are group together with a bootstrap value of 99%. These two sequences have a sequence identity of 63 % with each other but their similarity to the other sequences is less than 50%.They may represent another subgroup of elements which is underrepresented in this analysis. (Figure 3.1 and Figure 3.2A). Teq 29 groups with the elements of subgroup 3 with abootstrap value of 98% (Figure 3.1) but it has maximum nucleotide identity of 55% and maximum peptide sequence identity of 54% with the other members of Teq3

subgroup.So Teq 29 has been considered as an ungrouped sequence because of low sequence similarities to the rest of Teq3 subgroup elements. However it might be amemeber of another different but related subgroup of elements which is underrepresented in this piece of work.

3.2.7 %GC content of *Ty1 - copia* retrotransposons in *Agave tequilana*

In molecular biology the GC content is the percentage of nitrogenous bases on a DNA molecule which are either guanine or cytosine. This may refer to a specific fragment of DNA, RNA or that of the whole genome. The G (guanine) and C (cytosine) undergo a specific hydrogen bonding whereas A (adenine) makes specific bond with T (thymine). The level of GC contenet varies greatly among the various organisms (Mooers and Holmes 2000). Bacterial species have a wide range of GC content from 25% in *Microplasma capricolum* to 72% in the gram positive actinibacterium *Micrococcus luteus* (Deng et al., 1999). The percentage of GC content is a general parameter of the genome which reflects significant compositional features of the genome (Samarda et al., 2008). The GC content differes among different plant families (Barow and Meister 2002) however the GC the role of GC content in plant evolution especially in lower taxonomic groups is still unknown (Barow and Meister 2007). Recently many attempts have been made to see the role of GC content in the evulotion of plants. For example genome size and GC content in *Festuca* (Smadra et al., 2008) and the insertion and deletion with respect to GC content in *Arabidopsis thaliana* (Zhang et al., 2008).In this study the percentage GC content of different retrotransposon RT sequences was calculated and table 3.1 shows the GC content of all of the RT sequences isolated in this study.

Element	% GC content	Element	% GC content	Element	% GC content	Element	% GC content
Teq1	33.85	Teq12	43.89	Teq23	34.80	Teq34	34.36
Teq2	44.53	Teq13	46.42	Teq24	33.67	Teq35	34.62
Teq3	44.03	Teq14	45.43	Teq25	34.81	Teq36	33.27
Teq4	44.01	Teq15	45.54	Teq26	34.80	Teq37	34.19
Teq5	42.78	Teq16	45.52	Teq27	42.63	Teq38	35.14
Teq6	42.44	Teq17	44.10	Teq28	34.10	Teq39	35.23
Teq7	45.73	Teq18	45.74	Teq29	47.27	Teq40	43.61
Teq7	43.37	Teq19	46.56	Teq30	34.97	Teq41	32.38
Teq9	45.31	Teq20	42.89	Teq31	41.20	Teq42	35.23
Teq10	45.57	Teq21	42.78	Teq32	34.89		
Teq11	45.85	Teq22	36.55	Teq33	34.80		

Table 3.3 Percentage GC content of reverse transcriptase nucleotide sequences.

The percentage of GC content in the RT sequences of Ty1 copia retrotransposonss in *Agave tequilana* ranges from 33.27 in Teq36 to 47.27 in Teq29 as shown in the table 3.1.Tale 3.2 shows the average GC content of different subgroups of retrotransposon RT sequences. Teq1 and Teq24 subgroups have low percentage of GC content 34.49 % and 34.66% respectively while Teq2 and Teq3 contain a relatively high GC content with 44.36 in Teq2 and 43.26 in Teq3.

3.2.8 Comparison of Ty1- copia reverse transcriptase (RT) sequences with RT sequences from other organisms

Ty1-*copia* retrotransposon reverse transcriptase sequences isolated from *Agave tequilana* were compared to the other reverse transcriptase sequences isolated from different organisms in order to see them in the context of other well characterized Ty1-copia retrotransposons from plants and other organisms. Newly isolated RT sequences were used to perform blast searches using NCBI blast. The resulted sequences were manually sorted by looking at the conserved amino acids RLVAKG and YVDDM which flank the internal domain of reverse transcriptase. The collection of RT sequences from agave and the sequences from the data bases were aligned by Clustal W (Chenna et al., 2003). Phylogenetic trees were constructed and evolutionary relationship was determined MEGA 4 (Tamura et al., 2007). The trees and evolutionary histories were inferred using the Neighbour-Joining Method (Saitou and Nei, 1987)

Figure 3.6 shows the phylogenetic relationship among retrotransposon RT sequences from agave and other organisms. The details of the sequences obtained from NCBI blast is shown in the table 3.3. It can clearly be seen from Figure 3.6 that different subgroups of RT sequences from agave have similarities with RT sequences from different organisms.

A comparison with different well characterized retrotransposon families from other plants was done by getting the reverse transcriptase sequences of these retrotransposons from the data bases using NCBI blast. A phylogenetic tree of RT sequences from agave and the sequence obtained from the blast searches was constructed using MEGA4 (Molecular and Environmental Genetic Analysis) software. (Figure: 3.6 and 3.7) The accession numbers of the sequences used in the comparative analysis are shown in the table 3.3. A closer look at the phylogenetic tree in Figure 3.7 revealed the relationship among retrotransposon reverse transcriptase sequences identified in agave and the sequences of well characterized retrotransposons. Teq1 subgroup elements cluster together with *Ty1-copia* retrotransposon of *Oryza sativa* and the evolutionary relationship is supported by a high value (88%) of bootstrap confidence. On the other hand Teq24 elements seem to be related to BARE 1 retrotransposons of *Hordeum vulgare* while Teq3 and Teq2 elements do not show a significant relationship with retrotransposons used in this comparison.

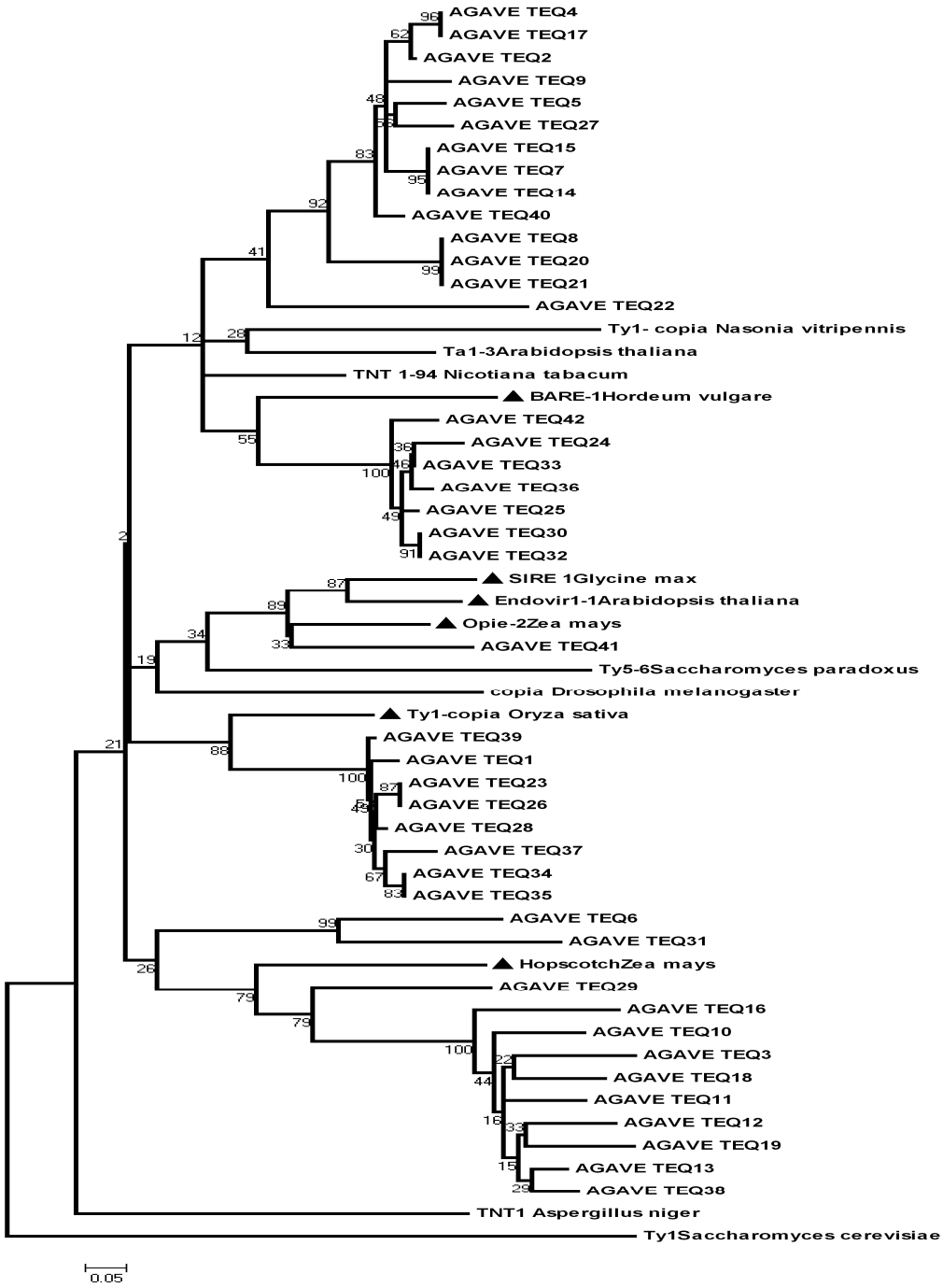


Figure: 3.6 Evolutionary relationships of Agave RT sequences with RT sequences from other retrotransposons. The evolutionary history was inferred using the Neighbor-Joining method. The bootstrap consensus tree inferred from 1000 replicates is taken to represent the evolutionary history of the sequences analyzed. Branches corresponding to partitions reproduced in less than 50% bootstrap replicates are collapsed. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (1000 replicates) is shown next to the branches. The tree is drawn to scale, with branch lengths in the same units as those of the evolutionary distances used to infer the phylogenetic tree. The evolutionary distances were computed using the Poisson correction method and are in the units of the number of amino acid substitutions per site. All positions containing gaps and missing data were eliminated from the dataset (Complete deletion option). Phylogenetic analyses were conducted in MEGA4. The retrotransposons grouping with RT sequences from *Agave tequilana* are shown with a black triangle.

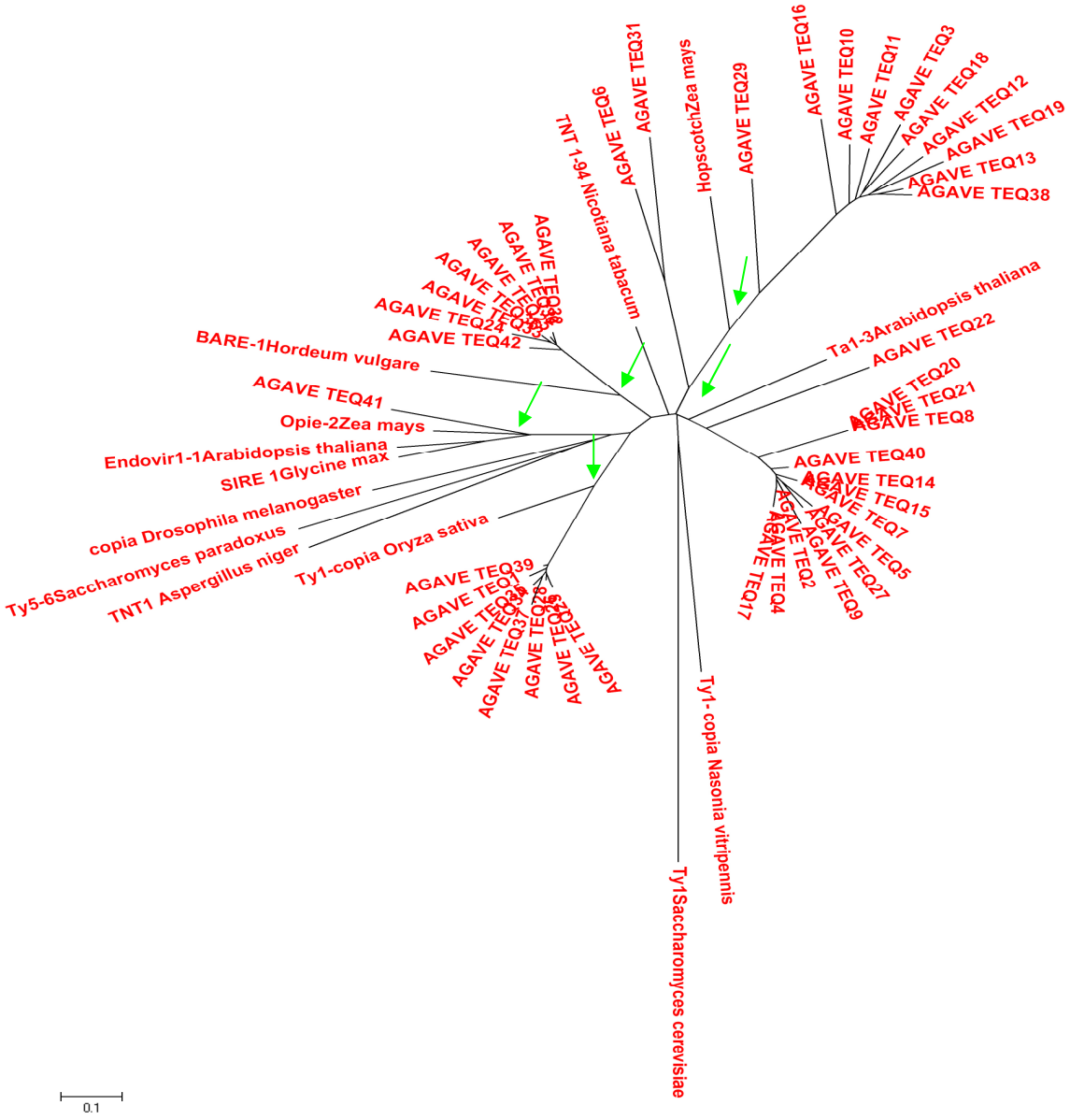


Figure:3.7 Evolutionary relationships of Agave RT sequences with RT sequences from other retrotransposons.

The phylogenetic relationship was inferred using the Neighbor-Joining method. The bootstrap consensus tree inferred from 1000 replicates is taken to represent the evolutionary history of the sequences analyzed. The tree is drawn to scale, with branch lengths in the same units as those of the evolutionary distances used to infer the phylogenetic tree. The evolutionary distances were computed using the Poisson correction method.. Phylogenetic analyses were conducted in MEGA4. The retrotransposons showing similarities and phlogenetic relationship with the agave retrotransposons are shown by green arrows

However Teq41 is related to Opie 2 of *Zea mays* and Teq29 shows evolutionary relation ship with Hospscotch retrotransposon of *Zea mays*. The comparison of previously identified well characterized retrotransposon reverse transcriptase sequences with RT sequence from agave reveald a diverse and heterogeneous nature of these retrotransposons in agave.

Name of the elemen	Accession Number	Name of the organism	Name of the elemen	Accession Number	Name of the organism
TNT 1-94	P10978	*Nicotiana tabacum*	copia	X04456	*Drosophila melanogaster*
TNT1	CAK44610	*Aspergillus niger*	Ty1	AAA66938	*Saccharomyces cerevisiae*
Ty1-copia	gi|30089754	*Oryza sativa*	Ta1-3	CAA31653	*Arabidopsis* thaliana
Ty1-copia	XP001599237	*Nasonia vitripennis*	Ty5-6	AAC02631	Saccharomyces paradoxus
SIRE 1	AAC64917	*Glycine max*	Hopscotch	AAA57005	*Zea mays*
Endovir1-1	AAG52949	*Arabidopsis thaliana*	BARE-1	Z17327	*Hordeum vulgare*
Opie-2	AAC49502	*Zea mays*			

Table 3.3 Ty1-copia type retrotransposons from different organisms used for comparison to retrotransposon RT sequences from agave. The sequences were obtained from the sequences databases using a blast search. The sequences were manually sorted by loking at the conserved motifs of RT domain and a phylogenetic analysis was performed comparing RT sequences from agave with them.

3.3 Discussion

3.3.1 *Agave tequilana* contains a population of heterogeneous *Ty1-copia* retrotransposons.

One of the aims of this study was to isolate a population of Ty1-copia retrotransposon reverse transcriptase (RT) fragments in order to characterize these retrotransposons and study their potential role in the evolution and organisation of *Agave tequilana* genome. The isolation methodology was based on the presence of conserved internal domain of retrotransposon reverse transcriptase flanked by two highly conserved motifs RLVA(KQ)G and YVDDM. The isolation was carried out by polymerase chain reaction (PCR) using degenerate primers for Ty1-copia reverse transcriptase as previously used in *Vicia* species (Pearce et al., 1996b), mungbean (Xiao et al., 2004), strawberry (Ma et al.,2008) and melon (Ramallo et al., 2008). The isolation experiment produced a population of 42 reverse transcriptase (RT) sequences deduced peptide sequences of which were found to be highly similar to RT sequences from other plants and the internal RT domain was highly conserved in all of the sequences. It is well established that segments within the RT domain are very well conserved and have been used to design degenerate primers for the isolation of *Ty1-copia* elements in a wide variety of plants like citrus (Tao et al., 2005), mungbean (Xiao et al., 2004) and sorghum (Muthukumar and Bennetzen, 2004).

The newly isolated population of RT sequences turned out to be highly heterogeneous with closely related sequences clustered together to make distinct groups of sequences. It is well established that the replicative transposition combined with the error prone nature of reverse transcription generates families of related sequences (Casacuberta et al., 1995) and resulting heterogeneity has been reported in many plant species (Flavell et al., 1992a; Flavell et al., 1992b; Friesen et al., 2001; Pearce et al., 1996a; Pearce et al., 1996b). So the isolation of *Ty1-copia* retrotransposons in agave generated a population of closely related heterogeneous reverse transcriptase (RT) sequences giving rise to four major subgroups of elements namely Teq1, Teq2, Teq3, and Teq24 subgroup revealed by the phylogenetic analysis of the sequences.

3.3.2 Major Subgroups of *Ty1-copia* retrotransposons are heterogeneous, abundant and potentially active in *Agave tequilana*

The isolation of retrotransposon reverse transcriptase (RT) fragments revealed the presence of four major subgroups of elements and some ungrouped sequences as described above. A closer look at the characteristics of all of these groups suggests that they can possibly be high copy number and potentially active retrotransposons.

The characteristics of Teq1 subgroup elements like shorter branch length for most of its member elements suggests a short evolutionary relationship among its elements, except for Teq37 which had longer branch length as compared to the rest of the group. Highly conserved internal RT domain and fully intact coding domain with an open reading frame with no frameshifts or stop codons coupled with the frequent occurrence of Teq1 elements not only points towards a recent activity of these elements but also suggests a potential future activity of Teq1 elements in agave. These characteristics also indicate a potential abundance of these elements in the genome of *Agave tequilana*. This type of groups of closely related elements were later revealed to be copies of active elements in sweet potato (Tahara et al., 2004). On the basis of this data it could be concluded that Teq1 subgroup represents a high copy number recently replicated *Ty1- Copia* retrotranspososn that can potentially comprise a substantial proportion of *Agave tequilana* genome.

Like Teq1 subgroup Teq2 subgroup is also a major subgroup of *Ty1-copia* elements in agave. The elements of this group are much diverse than Teq1 elements as the branch length of different members of this subgroup is different for different elements. As far as the integrity of the sequences is concerned 70 % of the Teq2 elements are uninterrupted and 30% have stop codons in their coding domain. Teq2 is the largest subgroup of sequences with high sequence integrity, highly conserved internal reverse transcriptase domain, high bootstrap value on the phylogenetic tree (88%). All of these characteristics make Teq2 a potentially high copy number, recently active subgroup of retrotransposon which can make a major proportion of agave genome.

The phylogenetic analysis of Teq3 revealed that it is different from other subgroups. Unlike Teq1 and Teq2 elements the members of Teq3 subgroup have longer branches showing more evolutionary distances among members of this subgroup. The sequence analysis of Teq3 elements showed that more than 70% reverse transcriptase sequences of this subgroup have either a stop codon or a frameshift or both of them. The presence of stop codons and frameshifts is not uncommon in LTR retrotransposons and paradoxically, for most of the LTR retrotransposons that have been described in the literature, most of the copies retrieved from the genomic sequences have been shown to be defective due to stop codons, frameshifts , insertions or deletions (Vitte and Panaud, 2005). Additionally many copies of defective retrotransposons with frameshifts and in frame stop codons may actually be the offspring of functional retrotransposons that have acquired these mutations due to the error prone nature of the reverse transcription (Keulen et al. 1997; (Gabriel et al., 1996) or these mutation may infect be post transpositional mutations (Wicker and Keller, 2007).

Although most of the Teq3 elements can possibly be the offspring of functional element but the presence of frameshifts and stop codons make them defective and incapable of replication. Moreover the longer branch length for almost all of Teq3 subgroup elements in the phylogenetic tree also suggests that these elements might not have been active in the recent past. However their occurrence as a cluster of closely related sequence and high bootstrap value in phylogenetic tree (99%) make them a distinct group of elements which might have been actively producing copies in distant past most of which have acquired mutations and are no longer capable of replication. But retrotransposons can also occur in the form of nonautonomous elements that are partially or fully dependent on the proteins expressed elsewhere in the genome (Vitte and Panaud, 2005). For example BARE 1 element of barley has a defective gag domain and probably use the protein coded by closely related BARE 2 elements for their replication (Sabot and Schulman, 2006). Nonautonomous retrotransposons with highly degenerated coding regions or no coding regions are commonly found in all classes of retrotransposons (Kelendar et al., 2004; Bureau and Wessler, 1994; Wicker et al., 2003b). On the basis of the characteristics of Teq3 elements it can be concluded that most of these elements might not be capable of replication and transposition. However they formed a major subgroup of elements with a high bootstrap confidence value suggesting that they might still be an abundant family of elements in the genome of agave. In addition their Long Branch length indicates no significant activity in the recent past but they can possibly occupy a large fraction of the host genome in the form of defective copies of *Ty1-copia* retrotranspososns.

In addition to above mentioned subgroups Teq24 is another major subgroup consisting of seven closely related reverse transcriptase sequences. The characteristics of Teq24 elements like sequences integrity, short branch length on the phylogenetic tree, and high bootstrap confidence value make them highly conserved, abundant, and potentially active *Ty1- copia* retrotransposons. Majority of Teq24 sequences isolated in this study have a short branch length showing a closer evolutionary relationship among them. Additionally they have highly conserved coding domain of their reverse transcriptase with 100% sequence integrity as all of the sequences have no stop codon or frameshifts. Moreover high bootstrap confidence value (100%) shows that they are a distinct subgroup of sequences and potentially a distinct family of *Ty1-copia* retrotransposons. All of these characteristics suggest that Teq24 is a subgroup of retrotransposons which might have been active in the recent past. Same type of closely related elements grouping together were later found to be active in sweet potato (Tahara et al., 2004). In addition to the major subgroups there is a small group of three sequences (Teq8, Teq20, and Teq21) which are also highly conserved but two of them have frameshifts in their coding domain making then defective.

Teq6 contains in frame stop codon and teq31 possesses a frameshift in its coding reverse transcriptase domain while Teq29 contains both of them. Teq22 and Teq41 have very long branches on the phylogenetic tree and with the other elements. However both of these elements are highly intact without any stop codon or frameshift. This means that there may be other families or subgroups of Ty1-copia elements in agave that could not be isolated in this study.

Collectively speaking three of the four major subgroups of elements isolated during this study are possibly high copy number and heterogeneous retrotransposons. These elements are possibly the offspring of recently active elements and potentially active themselves. These subgroups of elements and other ungrouped elements might have played an important role in the evolution and organization of *Agave tequilana* genome. Although there may be other families of Ty1-copia elements in the genome of agave and elements isolated in this study may not be able to portray the actual picture of the genome of agave in terms of retrotransposons, nevertheless the bulk of the elements discussed here may occupy a major fraction of *Agave tequilana* genome. It is now widely accepted that retrotransposoition have plyed a crucial role in the genomic expansion and architecture and could also have an impact on the regulation of gens in the major crop species (Kahskush et al., 2003). A recent study revealed that the genome size of *Oryza australiensis* a wild relative of cultivated rice *Oryza sativa* has doubled in last three million years due to the accumulation of three families of LTR retrotransposons (Piegu et al., 2006) indicating the dramatic impact of retrotransposons on the genome reshaping of plants. As retrotransposons are capable of genome organization it could be speculated that retrotransposons might also have played a vital role in the evolution of agave genome.

GC content has also been suggested as a general perameter of the genome and many studies in humans and other organisms revealed that GC content may have a significant compositional role in the genomes (Smarda et al., 2008). Recently measurements of GC content and its possible link with the genome size have been reported in *Festuca.* However it is assumed that the differeneces in the GC contents of different species has connections with the retrotransposons. For example the geome of *Oryza australiensis* which is a wild relative of *Oryza sativa* has doubled in last three million years due to the massive amplification of three retrotransposons, RIRE1, Kangourou and Wallabi (Piegu et al., 2006). The GC content of these three elements is 44.6%, 50% and 50.9% respectively, which is considerably more than the average GC content of the genome (43.6%) and even greater than the GC content of the related *Oryza sativa* (45.3%) genome (International Rice Genome Sequencing Project, 2005). The assumed GC content of the O. Australiensis was recently confirmed by the partial sequencing of its genome (Ammiraju et al., 2006). It is suggested that the retrotransposon driven genome expansion of *O. australiensis*

may be analogous to the fescues and the GC content expansion and reverse reduction (Smarda et al., 2008) may reflect a long term dynamics of GC rich retrotransposon proliferation and removal. The GC content of fescues and related genera is considered among high GC contents in angiosperms, ranging from 42.5 % to 46.4% (Smardaet al., 2008) while GC content of most of the elements studied here is over 40%. The GC content of *Ty1-copia* retrotransposons analysed in this chapter might also have an impact on the evolution of agave genome.

3.3.3: *Ty1-copia* retrotransposons from *Agave tequilana* showed evolutionary relationship with well characterized retrotransposons from other species.

Comparison of isolated *Ty1-copia* retrotransposon with well characterized *Copia* type retrotransposons from other species is a well established tool used to see the evolutionary relationship among different families of retrotransposons. This type of comparison has previously been used in mung bean (Xiao et al., 2004) and melon (Ramallo et al., 2008). Comparison of different subgroups of retrotransposon RT sequences from agave as well as individual ungrouped sequences with RT sequences from other species revealed an evolutionary relationship among them.

Two major subgroups namely Teq24 subgroup and Teq1 subgroup have clearly shown a phylogenetic relationship with *BARE1* elements of *Hordeum vulgare* and *Ty1- copia* elements of *Oryza sativa* respectively. On the other hand Teq41 which is an individual ungrouped element was found to be related to *Opie1* of *Zea mays, Sire*1 of *Glycine max* and *endovir*1-1 of *Arabidopsis thaliana.* Similarly Teq29 shows some relation ship with *Hopscoch* elements of *Zea myse.* This type of comparison not only shows the diversity of Ty1-copia elements in agave but also points out the common evolutionary history of all major families of Ty1-copia elements. It is interesting that the main evolutionary lineages and sublineages of copia elements are conserved between species that diverged 50 million years ago such as rice and Triticeae or 140-150 million years ago (monocot and dicot) (Paterson et al., 2004 ; Chaw et al., 2004). The genome sequencing of rice and *Arabidopsis* prvided an opportunity to compare retrotransposons from different plant species and recently it was reported that most of copia elements from Triticeae have homologs in the rice genome; However less homology between *Arabidopsis* and Triticeae elements was found (Wicker and Keller, 2007). The similarities between the retrotransposons of agave and other plants like *Arabidopsis*, rice, and barley found in this study indicated that the retrotransposons of agave might predate the divergence of monocots and dicots and also suggests that a horizontal transfer between these plant species might have occurred. The *Adena* elements of rice also showed strong sequence conservation in rice and the possibility of a horizontal

transfer between these two species was suggested (Wicker and Keller, 2007). A detailed comparative analysis of retrotransposons from agave with rice and *Arabidopsis* retrotransposon might give a further insight into the common origin of retrotranspsons in these species, however this piece of work could be the basis of detailed analysis of different retrotransposons in agave and their contribution in the evolution and organization of agave genome and such an analysis might help in solving many issues surrounding this economically important crop.

In conclusion *Agave tequilana* contains a diverse population of *Ty1-copia* retrotransposons which occur in the form of different subgroups of elements. These subgroups of elements are related to different well characterized elements from other species of plants as well as other organisms.

CHAPTER 4

Ty1- *copia* retrotransposon copy number in *Agave tequilana*

4.1 Introduction

The blue agave, *Agave tequilana* Weber var. Azul is the only variety legally permitted for the production of tequila by Mexican government and has been vegitatively propagated through clonal offshoots for the last 200 years. The blue agave is assumed to have a very low genetic diversity due to clonal propagation and legal restriction on the plantation. This narrow genetic diversity has led to the suggestion that existing agave plantation may be vulnerable to the diseases, pathogens and adverse environmental conditions (Vega et al., 2006). Although blue agave is a little understood plant but the growing concern about the genetic diversity and the economic importance of the plant has led to an increased interest in genetic and biochemical research in recent years (Dalton, 2005).

Retrotransposons are the most abundant and commonest class of eukaryotic transposable elements (Bennetzen, 1996; Grandbastien, 1992; Kumar, 1996; Kumar and Bennetzen, 1999; Pearce et al., 1996b). They are the largest group of transposable elements and are the main constituents of the large plant genomes(Kumar and Bennetzen, 1999). They are widespread in plant genomes and are considered to have an important role in genome evolution by causing mutations, reorganizing, and contributing to the physical size of the genomes (Kidwill and Lisch 1997, Kumar and Bennetzen 1999). Retrotransposons transpose via an RNA intermediate which is reverse transcribed before integration into a new location in the genome (Gradbastien, 1992; Kumar and Bennetzen, 1999). Retrotransposons have been to be present in a high copy number in heterochromatic regions including centromeres (*Arabidopsis* Genome initiative 2000; Feng et al., 2002) but they are also found interspersed with genes, for example retrotransposon like sequences have been found in the regions flanking maize zein genes (White et al., 1994). Retrotransposons can give rise to nested structures by inserting into pre existing retrotransposons like adh1 region in maize (SanMiguel et al., 1996). The differences in the genome sizes of different plant species like maize and sorghum (SanMiguel et al., 1998) and tow species of *Oryza* genus namely *Oryza sativa* and *Oryza australiensis*(Piegu et al., 2006; Vicient and Schulman, 2002) have been attributed to the contribution of retrotransposons.

LTR retrotransposons are an important source of genetic diversity and have had a major impact on the structure of plant genomes (Kumar and Bennetzen, 1999; Pearce et al., 1996b; SanMiguel et al., 1996). Due to their replicative mode of transposition LTR retrotransposons can successfully amplify and accumulate in high numbers in plants, which can often lead to a significant expansion of plant genome size (Bennetzen and Kellogg, 1997). Plant species with large genomes like maize and barley tend to contain high copy number retrotransposon families (Kumar and Bennetzen, 1999), whilst this type of elements are usually underrepresented in plants with small genomes, like rice and *Arabidopsis* (McCarthy et al., 2002; Pereira, 2004).

LTR retrotransposons presumably multiplied in the course of host plant evolution (Tahara et al., 2004). They can account for large proportions of their host genomes for example they occupy 50-80% of the maize (Sanmiguel and Bennetzen, 1998), 70% of barley (Vicient et al., 1999b), 20% of rice (Takata et al., 2007), 26% of melon (Ramallo et al., 2008), 15.8% of cultivated strawberry (Ma et al., 2008), 23% of the citrus (Rico-Cabanas and Martinez-Izquierdo, 2007) and 5.6% of *Arabidopsis* genome (Pereira, 2004).

Ty1- *copia* retrotransposons are the best studied LTR retrotransposons in plants. They are present in high copy number populations of heterogeneous sequences and are distributed through out the genomes of plants (Heslop-Harrison et al., 1997; Kumar et al., 1997; Pearce et al., 2000). In some species of plants Ty1- *copia* group elements are so numerous that they comprise major fractions of the genome (Gribbon et al., 1999; Konieczny et al., 1991; Todorovska, 2007). The copy number of *copia* type elements in plants varies over four orders of magnitude, from several hundred elements in *Arabidopsis thaliana* to around one million in *Vicia faba* (Pearce et al., 1996b). *Arabidopsis thaliana* has a small genome (1C= 0.15 pg) which yields a total of 779 elements (Schulman et al.,2004) with 310 *copia* elements (Navarro-Quezada and Schoen, 2002). The genome of rice *Oryza sativa* contains a population of 1000 retrotransposons of which 100 elements belong to *Ty1-copia* group (Wang et al., 1999). Ty1-copia elements are also present in high copy number in the members of *Vicia* genus however the copy number of Ty1-copia retrotransposonss is very different in different vicia species with 1000 copies in *Vicia melanops* ,5000 copies in *Vicia sativa* and 30,000- 1 million (10^6) copies in *Vicia faba* (Pearce et al., 1996b). The differences in the genome size of grass species is attributed to the presence of retotransposons as they generate nested clusters surrounded by gene dense regions significantly increasing the size of genomes (Bennett and Leitch, 2004). The copy number of these elements is very different in individual grass species, for example the genome of barley *Hordeum vulgae* contains approximately 196,000 copies while *Triticum aestivum* has 80,000 copies of *copia* type elements (Liu and Somerville., 1996).

Ty1-copia elements are also numerous in *Pisum* and *Picea* species and their copy number is different in different species for example there are approximately 4000 copies of Ty1-copia retrotransposons in *Pisum sativum* (Pearce et al., 2000) and 0.5-1x 10^6 copies in *Picea abies* (Pearce et al., 2000). The genome of *Citrus sinensis* contains 9900 *copia* type elements along with 4900 *gypsy* elements (Rico-Cabanas and Martinez-Izquierdo, 2007) the genome of cultivated strawberry (*Fragaria ananassa*) contains 2800 copies (Ma et al,.2008) and there are 6,800 *copia* elements in melon (*Cucumis melo*) genome (Ramallo et al., 2008).
The *Agave tequilana* genome (2C = 8.8pg or 8624Mbp) is larger than the genome of barley (2C = 5.45pg) and smaller than that of of maize (2C = 11.1pg), but it is significantly larger than the small genomes of rice (2C = 1.00pg) and *Arabidopsis* (2C = 0.32pg) (Bennetzen, 2000; Sabot et al., 2005; Sanmiguel and Bennetzen, 1998). They comprise 50- 80% (Meyers et al., 2001; Sanmiguel and Bennetzen, 1998) of maize and 70% of barley genome (Feschotte et al., 2002; Vicient et al., 1999b) while they contribute to 20% of rice (Takata et al., 2007) and 5.6 % of *Arabidopsis* genomes (Pereira, 2004).

In the previous chapter it was observed that *A. tequilana* has at least four major subgroups of Ty1-*copia* retrotransposons. The presence of clusters of closely related elements indicates that these particular elements may have been actively replicating and producing new copies in the recent past. The sequence data although giving a rough overview of the overall population structure of Ty1- retrotransposons does not give any information on the numbers of these elements in the *A. tequilana* genome. In order to estimate the abundance of *Ty1-copia* elements and their contribution to the genome of *Agave tequilana* a detailed copy number estimation study was conducted using quantitative slot blotting and hybridization techniques. One of the aims of copy number estimation by quantitative slot blot was to confirm that the information provided by the sequence and phylogenetic analysis as well as confirmation of isolation tree. The quantitative slot blot would show the abundance and heterogeneity of major subgroups of elements as well as ungrouped elements. Some of the ungrouped elements may actually be separate subgroups themselves but underrepresented in this analysis. Copy number estimation and evaluation of heterogeneity would give an idea about their real status in the genome of *Agave tequilana*.

4.2: Use of Quantitative slot blotting for the estimation of retrotransposon copy number and heterogeneity of retrotransposons.

Slot blot analysis is a technique used for the estimation of copy number and heterogeneity of sequences. The basic principals of slot blot are same as conventional Southern blotting or in other

words slot blotting is a quantitative Southern blot. In slot blotting the copy number is estimated by the comparison of the intensity of hybridization signals between control DNA and Genomic DNA.

Precise amounts of control DNA containing the individual probe sequences are loaded in one row of the slots and precise amounts of genomic DNA are loaded in the other row of slots. The probes hybridise to the serial dilutions of genomic DNA as well as a serial dilutions of themselves to control for probing efficiency. The hybridization signals of the probes to themselves (control DNA) are compared with the hybridization signals of the probe to the genomic DNA. As the amount of DNA loaded to each slot is accurately measured the comparison of the hybridization signal can give a good estimate of the element copy number in the genome (Pearce et al., 1996b). A series of stringent washes is performed on the blot to evaluate the copy number and heterogeneity of the sequences. For example at the lowest stringency (55°C in 1x SSC) the probe will hybridise to all of the related sequences containing a sequence identity of 70%- 100%. At a higher stringency (60°C in 1x SSC) the sequences with less sequence identity will be washed off and only the sequences with more than 80% sequence identity will be hybridized and at a stringency even higher (65°C in 1x SSC) only closely related sequences with sequence identity of more than 90% will hybridise. At the highest stringency (65°C in 0.1x SSC) only the sequences with absolute sequence identity and sequences with identity close to 100% will be hybridized. Thus different stringent washes would wash the sequences away showing the heterogeneity of sequences.

Slot blot analysis has been utilized in several plants for the determination of retroelement copy numbers, including *Vicia* and *Avena* species (Linares et al., 1999; Pearce et al., 1996b), rye (Francki, 2001; Pearce et al., 1997), onion (Pearce et al., 1996c), oat (Linares et al., 2001) and olive (Natali et al., 2007). Usually, the probe consists of a cloned reverse transcriptase PCR fragment that can be amplified with primers between RT conserved domains and is specific for each family. To estimate the copy number of retrotransposons in the genome of blue agave, cloned reverse transcriptase (RT) sequences isolated by using degenerate RT primers in chapter 3 were used as probes.

4.3: Results

To estimate the copy number of *Ty1- copia* retrotransposons in *Agave tequilana* genome cloned reverse transcriptase (RT) fragments representing each subgroup of sequences (Figure: 4.1) were selected to be used as probes. The genomic DNA and control DNA were quantified and diluted

to make stock solutions. As the genome size of blue agave is known (8.8pg) the number of genomes in a given amount of DNA can be calculated.

The molecular weight of 1pg DNA is 3.9×10^{11} g while molecular weight of 8.8 pg(*Agave tequilana*) genome is 2.6×10^{12} g (3.9×10^{11} x 8.8). Now the number of genomes in 0.1ug DNA of *A.tequilana* can be calculated by the following formula

$$\frac{\text{(Avogadro's number) (0.1ug DNA expressed in grammes)}}{\text{Molecular weight of A.tequilana genome}}$$

Number of genomes in 0.1µg of *Agave tequilana* DNA= $\frac{(6.02 \times 10^{23})\,(1 \times 10^{-7})}{2.6 \times 10^{12}} = 2.3 \times 10^{4}$

0.1 ug agave genomic DNA was serially diluted by a factor of 10 to be loaded in the subsequent slots loaded in one row of slots. On the other hand for the control DNA to represent a single copy of the element same number of target molecules need to be loaded as the number of genomes loaded above (2.3×10^4genomes). The actual weight of 2.3×10^4 molecules of a control DNA of a given length can be calculated. If the length of a fragment of DNA is 100bp the weight of 2.3×10^4molecules of this DNA will be calculated as

$$\frac{(310 \times 100)(2.3 \times 10^{4})}{6.02 \times 10^{23}} = 1.18 \times 10^{-15}\text{g} = 1.18 \times 10^{-9}\ \mu\text{g}$$

In this way the weight of control DNA was calculated and precise amount of each control DNA was loaded in the respective slots. So the dilution of control slot would contain 2.3×10^4 molecules of control DNA and it will represent one copy of the element. The number of copies would increase by a factor of 10 in the subsequent slots reaching up to 1×10^{10} copies in the last slot (Figure 4.2). Probes specific to each subgroup were hybridized to the genomic DNA as well as control DNA and the copy number was estimated by comparing the hybridization signals of genomic DNA slots and the control DNA slots. The decrease in the intensity of hybridization signals would show the heterogeneity that a particular subgroup of sequences has in the genome of *Agave tequilana* (2.12 in general materials and methods, chapter 2)

Figure 4.1 shows different subgroups as well as ungrouped cloned reverse transcriptase (RT) sequences. To estimate the copy number of retrotransposons of each subgroup, one cloned RT sequences was selected from each subgroup along with five individual ungrouped sequences as shown in Figure: 4.1.each probe was then used for a detailed copy number estimation and

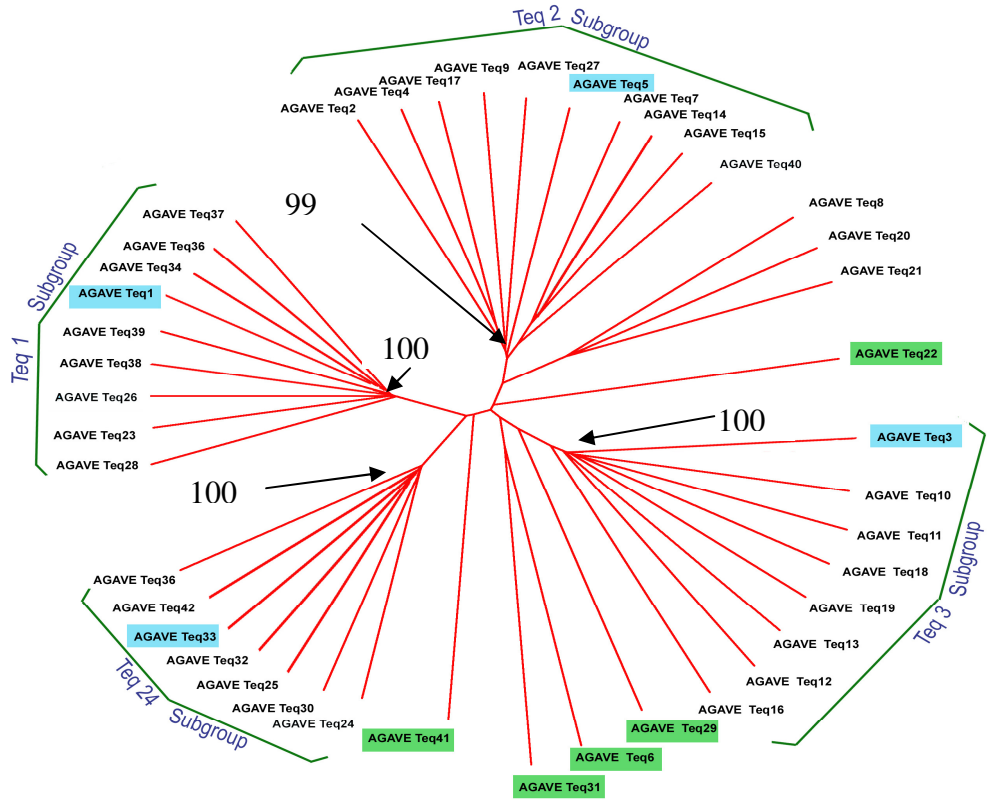


Figure 4.1: Unrooted tree of reverse transcriptase (RT) nucleotide sequences.
The tree is based on multiple sequence alignment of reverse transcriptase nucleotide sequences from *Agave tequilana* using CLUSTAL W. Four major subgroups are shown by green brackets. The sequences used as probes in slot blot analysis for the subgroups are shaded in light blue and the ungrouped sequences used as probes are shaded in green

heterogeneity studies. Different stringent washes were carried out on the membranes. The lowest stringency (55°C in 1x SSC) generates a probe target that allows all of the copies of the elements which belong to that subgroup to be identified, giving us the total copy number of the subgroup. As the stringency goes higher (60°C in 1x SSC, 60°C in 0.1x SSC, 65°C in 0.1x SSC), the number of target molecules for the probe decreases allowing only very closely related elements to hybridise and to be identified. These stringencies can give an estimate of the heterogeneity of the elements.

4.3.1: Copy number estimation of Teq1 subgroup

Teq1 is one of the major subgroups of subcloned RT sequences from *Agave tequilana* with a minimum nucleotide similarity of 94%. The two most different nucleotide sequences are Teq23 and Teq37 as they are 6% different from each other. So the maximum heterogeneity among the Teq1 subgroup elements is 6%. The copy number of this subgroup was estimated by using reverse transcriptase sequences Teq1 which has 95% sequence similarity to the rest of the subgroup.

For Teq1 (Figure: 4.2) the intensity of the hybridization signal produced at lowest stringency (55°C in 1x SSC) by slot 6 with agave genomic DNA that contains 2.3 $x10^4$ genomes (Row G, slot 6 picture a) is approximately equal to the intensity of the hybridization signal in slot 5 of the control DNA lane which contains 2.3 $x10^9$ target molecules (10^5 more copies). This result shows that there are approximately 100,000 copies of Teq1 subgroup elements in the genome of *Agave tequilana*. By washing the same hybridization at a series of increasing temperatures it was possible to gain data on the level of sequence heterogeneity within the subroup. If the Teq1 subtroup contained a large number of more diverse sequences then with increasing stringency the signal would drop. If on the other hand the population was made up of elements which were all very similar to the probe, the signal would remain high as the wash temperature increased. A series of stringency washes was carried out on the membrane as shown in b (60°C in 1x SSC), c (65°C in 1x SSC) and d (65°C in 0.1x SSC) of Figure 4.2. At the lowest stringency all of the Teq1 subgroup elements would hyberidize to the probe producing a high intensity signal and the signal would drop down with an increase in the stringency as more and more elements which are different from the probe would be washed away. It is clear from the Figure 4.1 that signal intensity is high at the low stringency (55 °C and 1x SSC) (Panel a Figure 4.2), it remains constant at higher stringency (60°C in 1xSSC) but it decreases slightly at highest stringency (65°C in 1x SSC) as shown in the panel c Figure 4.2 . The highest stringent wash in the weakest

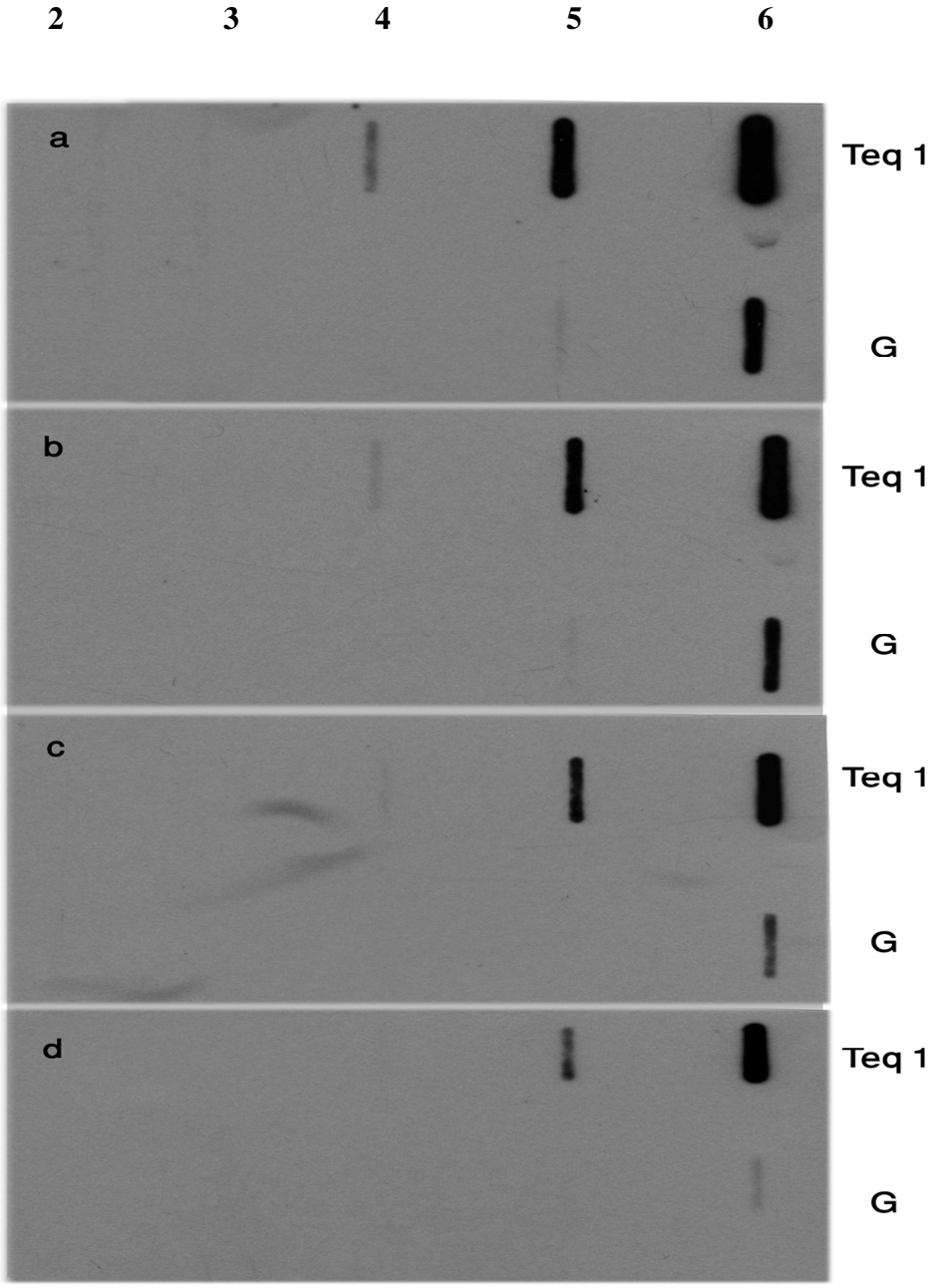


Figure: 4.2 Quantitative slot blot of Teq1 subgroup.
Cloned reverse transcriptase fragment derived from Teq1 ($2.3x10^{10}$molecules in slot 6and $2.3.x10^{9}$ molecules in slot 5 and $2.3.x10^{8}$ molecules) were fixed to the membrane. Row G contains genomic DNA from *Agave tequilana (*slot 6 contains 2.3 $x10^{4}$ genomes). Blots were probed and hybridized with cloned reverse transcriptase fragment of Teq 1and different stringent washes were carried out on the membrane. Panel "*a"* shows the lowest stringency (55°C in 1x SSC), "*b"* and "*c"* represent higher stringencies (60°C in 1x SSC and 65°C in 1xSSC) while "*d"* shows the highest stringency (65°C in 0.1xSSC)

buffer however decreases the signal intensity (Panel d Figure 4.2). However the signal intensity could also be affected by the Tm of the probe. The Tm of Teq1 is 77°C (Table 4.1) which is higher than 65°C but still low as compared to the other probes (Table 4.2). So a low Tm could also be a reason for the drop in the signal intensity of Teq1 at 65°C in 0.1x SSC. From the results describe above we can conclude that there are approximately 100,000 copies of Teq1 subgroup elements in the genome of *Agave tequilana,* and Teq1 subgroup is heterogeneous as well as high copy number.

4.3.2: Copy number estimation of Teq2 subgroup

Teq2 subgroup is the biggest subgroup of subcloned reverse transcriptase (RT) sequences consisting of elements. According to the sequence data Teq5 subgroup can have a maximum of 21% heterogeneity as Teq40 has a nucleotide sequence identity of 79% with Teq5 and these two sequences are the most different sequences in the subgroup. So Teq40 is the most divergent sequence of Teq2 subgroup.

For Teq5 (Figure 4.3) the intensity of the hybridization signal produced by the 2.3 $x10^4$ agave genomes (row G slot 6) is approximately equal to the hybridization signal produced by 2.3 $x10^9$ target molecules of Teq5 control DNA (control slot 5). So the copy number of Teq5 type elements is approximately about 10,000 copies in the genome of *Agave tequilana.* As shown in the Figure 4.3 the hybridization signal reduces with the subsequent washes becoming very low at the highest stringency (65°C in 0.1x SSC).

The subsequent stringent washes (Panels b, c, and d in Figure 4.3) resulted in a reduction in the intensity of the hybridization signal which means that Teq2 subgroup is heterogeneous in nature. Taq2 subgroup is a diverse group of sequences as the nucleotide sequence identity varies between 79% (Teq5 and Teq40) to 100% (Teq7 and Teq14). It can be seen from Figure 4.3 that the hybridization signal drops sharply at the highest stringency (65°C in 0.1x SSC) as at this stringency the probe hybridizes to itself and all the other elements are washed away by the increasing wash temperature and decreasing buffer strength.

As a whole Teq2 group is a diverse and heterogeneous subgroup of elements with a possible sequence heterogeneity of 21%, and is less numerous as compared to the other subgroups of elements. It is also possible that Teq40 is related to Teq2 subgroup but belongs to another subgroup of elements which are underrepresented here.

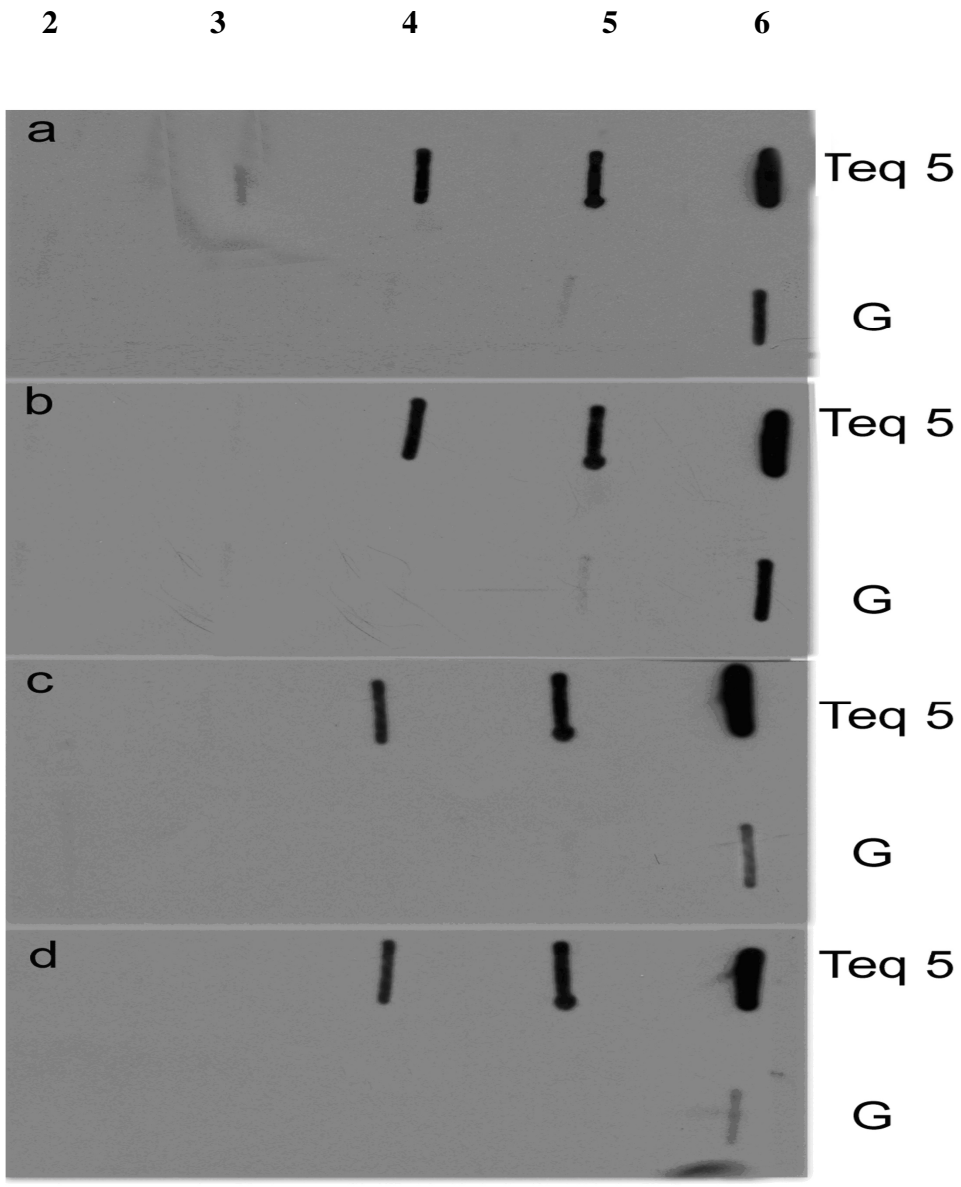


Figure: 4.3 Quantitative slot blot of Teq 5 subgroup
Cloned reverse transcriptase fragment derived from Teq5 ($2.3x10^{10}$molecules in slot 6 and $2.3.x10^{9}$ molecules in slot 5 and $2.3.x10^{8}$ molecules in slot 4) were fixed to the membrane. Row G contains genomic DNA from *Agave tequilana (*slot 6 contains 2.3 $x10^{4}$ genomes). Blots were probed and hybridized with cloned reverse transcriptase fragment of Teq 5 and different stringent washes were carried out on the membrane. Panel "*a*" shows the lowest stringency (55°C in 1x SSC), "*b*" and "*c*" represent higher stringencies (60°C in 1x SSC and 60°C in 0.1xSSC) while "*d*" shows the highest stringency (65°C in 0.1xSSC)

4.3.3: Copy number of Teq3 subgroup

Teq3 subgroup is another major subgroup of reverse transcriptase sequences. This subgroup is comprised of eight elements with a maximum nucleotide identity of 91% (Teq11 and Teq18). The most divergent sequence in this group is Teq16 which showed 82% nucleotide sequence identity with Teq12. So minimum heterogeneity in this subgroup is 9% and maximum sequence heterogeneity could go up to 18%. However most of the sequences in this grop have a sequence identity between 85% and 90%. Here in this group a maximum of 18% mismatch is possible. Figure 4.4 shows a quantitative slot blot of Teq3 subgroup and Panels a,b,c and d on Figure4.4 represent the change in the intensity of hyberidisation signal with increasing wash temperature.

The hybridization signal intensity of slot 6 (Figure: 4.4, panel a, row G) with by 2.3 $x10^4$ genomes coincides with the hybridization signal intensity of slot 5 of the control DNA lane containing 2.3 $x10^9$ target molecules of Teq3 control DNA. This comparison shows that there are approximately 100,000 copies of Teq3 subgroup elements in the genome of *A. tequilana.* To estimate the copy number and heterogeneity of this subgroup Teq3 was used as aprobe which shows minimum nucleotide sequence identity of 83% with Teq16 which is another member of the same group. The decrease in the hybeidisation signal intensity (Panel a,b,c,and d on Figure 4.4) indicates that Teq3 subgroup is heterogeneous. At the lowest stringency all of the Teq3 sequences would hyberidise to the probe and that means all of the elements in the genome of agave having sequence similarities of 90% or more would hyberidise to the probe, while at the highest stringency the only the probe would hyberidise to itselfe. Figure 4.4 panel d shows the reduced signal intensity and shows the individual copy number of Teq3 in the genome of agave. The Tm of the probe is 82°C which is well above the maximum wash temperature of 65°C so the drop in the signal intensity is due to the

On the basis of above results we can say that the genome of *Agave tequilana* contains approximately about 100,000 copies of Teq3 type elements and this subgroup is cluster of closely related high copy number elements.

4.3.4: Copy number of Teq24 subgroup

Teq24 is one of the major subgroups of subcloned RT sequences identified in the previous chapter which contains 7 elements..The nucleotide sequence identity among the members of this group is very high ranging from 92% to 98%. Teq42 is the most divergent sequence of this subgroup with 92% nucleotide identity with Teq 30. So Teq 42 and Teq30 are the most different

sequences in this subgroup. According to the sequence data the maximum heterogeneity among the members of this group is 8% while minimum heterogeneity would be 2%. Teq 33 was used as

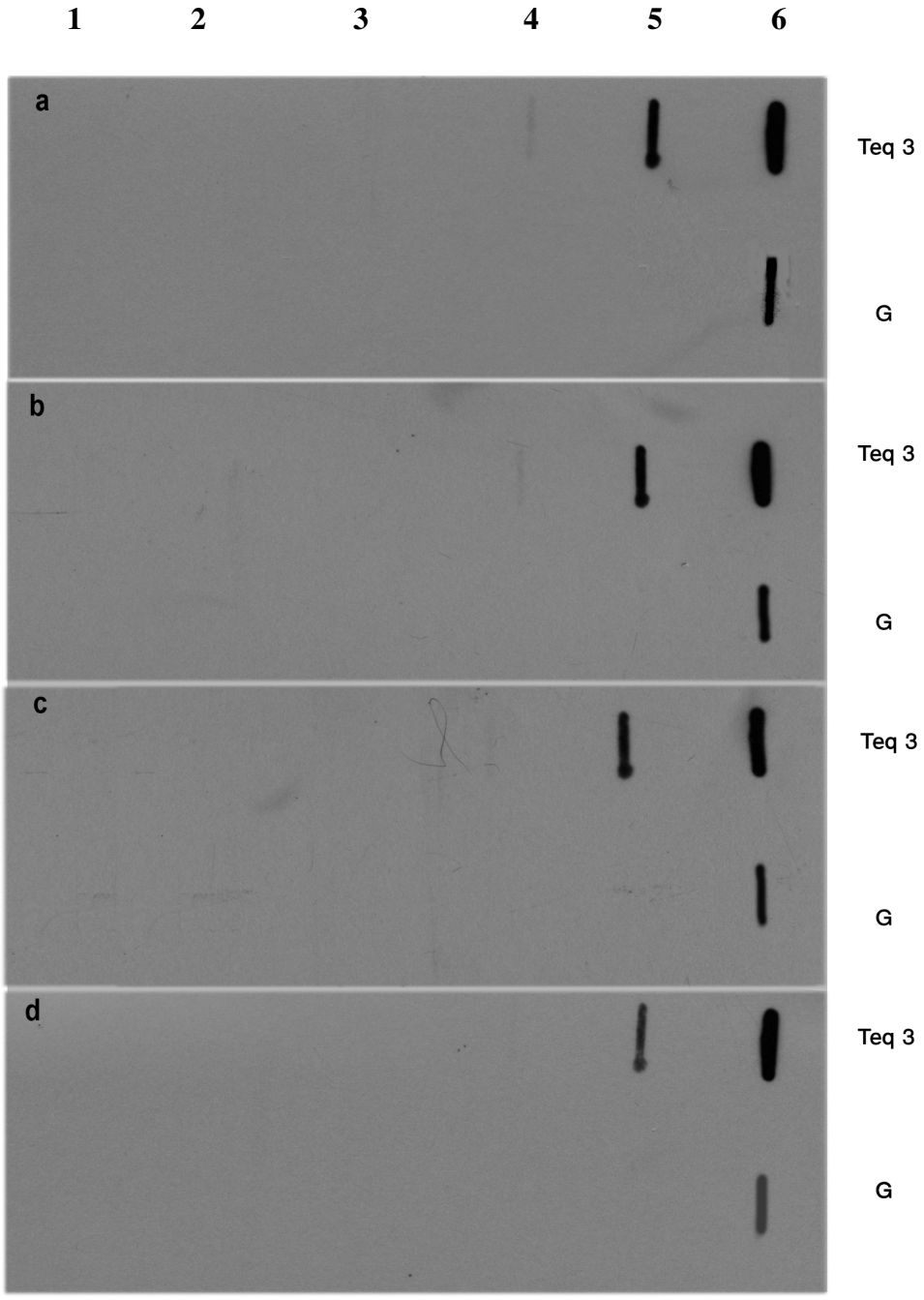


Figure: 4.4 Quantitative slot blot of Teq 3 subgroup
Teq 3 is a member of Teq 3 subgroup. Cloned reverse transcriptase fragment derived from Teq3 ($2.3x10^{10}$molecules in slot 6and $2.3.x10^{9}$molecules in slot 5, $2.3.x10^{8}$ molecules in slot 4) were fixed to the membrane. Row G contains genomic DNA from *Agave tequilana (*slot 6 contains 2.3 $x10^{4}$ genomes). Blots were probed and hybridized with cloned reverse transcriptase fragment of Teq 3 and different stringent washes were carried out on the membrane. Picture "*a*" shows the lowest stringency (55°C in 1x SSC), "*b*" and "*c*" represent higher stringencies (60°C in 1x SSC and 60°C in 0.1xSSC) while "*d*" shows the highest stringency (65°C in 0.1xSSC)

a probe in the quantitative slot blot to estimate the copy number and heterogeneity of the Teq24 subgroup.Teq33 has great similarities to the rest of the group. Figure 4.5 shows the copy number and sequence heterogeneity of Teq24 subgroup. It is clear from Figure 4.5 that the intensity of the hybridisation signal of 2.3 $x10^4$ genomes (Row G slot 6) coincides with the hybridization signal of 2.3 $x10^9$ target molecules of Teq33 control DNA (control slot 5 in panel a, Figure 4.5) which means that the total copy number of Teq24 subgroup is approximately about 100,000 copies in the genome of *Agave tequilana* because slot 5 in the control DNA lane contains 10^5 more copies.

The changes in the intensity of hybridization signal are shown by Panels a, b, c and d in Figure 4.5. However the intensity of hyberidisation signal did not decrease with the subsequent stringent washes in this case. However at the highest stringency (65°C in 0.1xSSC) the intensity of the signal decreased as only the probe itself and very closely related elements would hyberidise at the highest stringency. This result coincides with the sequence data as maximum heterogeneity was not expected to be more than 8%. As awhole Teq24 subgroup is a high copy number subgroup of retrotransposonss in the genome of *Agave tequilana* which consists of a heterogeneous population of elements.

4.3.5: Copy number of individual ungrouped elements

The multiple alignments and phylogenetic analysis carried out in the previous chapter (chapter 3) revealed that there are some individual sequences which do not group with any of the major subgroups on the phylogenetic trees. A detailed sequence analysis was carried out with respect to the nucleotide and peptide sequence identities as well as the phylogeny of these ungrouped sequences.Five of these elements were included in the quantitative slot blot analysis namely Teq6, Teq22, Teq29, Teq31 and Teq41.

The reason these ungrouped sequences were included in the analysis was to confirm that the isolation of retrotransposon RT sequences (Chapter3) and the tree based on these sequences was correct. The isolation through PCR and subcloning may amplify some elements but not others and phylogenetic trees based on these isolations may present some sequences as single sequences but actually they might be major subgroups underrepresented in a particular isolation study and of course some elements may not be there at all. So the estimation of the copy number of these single elements would present the actual picture of these elements in the genome of agave

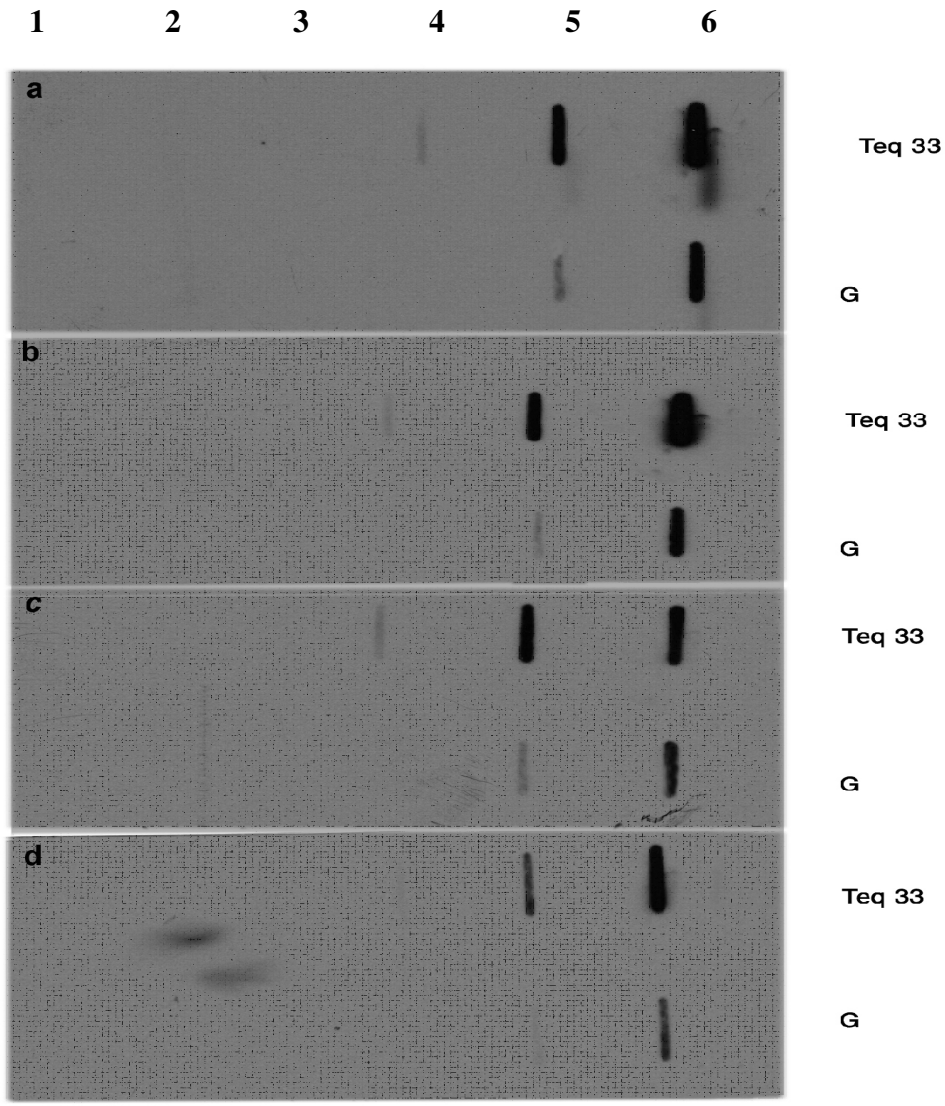


Figure: 4.5 Quantitative slot blot of Teq 24 subgroup
Teq 33 is a member of Teq 24 subgroup. Cloned reverse transcriptase fragment derived from Teq33 ($2.3x10^{10}$molecules in slot 6 and $2.3.x10^{9}$ molecules in slot 33 and $2.3.x10^{8}$ molecules in slot 4) were fixed to the membrane. Row G contains genomic DNA from *Agave tequilana (*slot 6 contains 2.3 $x10^{4}$ genomes). Blots were probed and hybridized with cloned reverse transcriptase fragment of Teq 33 and different stringent washes were carried out on the membrane. Panel "*a"* shows the lowest stringency (55°C in 1x SSC), "*b"* and "*c"* represent higher stringencies (60°C in 1x SSC and 60°C in 0.1xSSC) while "*d*" shows the highest stringency (65°C in 0.1xSSC)]

Teq6 is one of the ungrouped sequences and is not significantly related to any of the subgroups. However it has a nucleotide sequence identity of 71% and peptide sequence identity of 63% with Teq31. Teq6 and Teq31 also appear together on the phylogeneytic tree with a bootstrap value of 99%. Apart from Teq31 the sequence identity of Teq6 with the rest of the sequences in the population is not more than 50%. For Teq6 (Figure: 4.6) the intensity of hybridization signal of 2.3 $x10^4$ genomes (Row G slot6) is approximately equal to the hybridisation signal in control slot 4 which contains 2.3 $x10^8$ target molecules of teq6 control DNA. A careful comparison of the slots shows that the copy number of Teq6 elements in the genome of agave is between 10,000 and12,000 as the hybridization signal of genome slot (2.3 $x10^4$ genomes) is slightly stronger than control slot (2.3 $x10^8$ target molecules). The intensity of the hybridization signal decreases as a result of stringent washes and illustrated in the subsequent panels (panels a, b, c, and d) of Figure 4.6. The reduction in the hybridization signal with the stringent washes shows that there are other elements similar to Teq6 in the genome of agave which were washed away with an increase in the wash temperature indicating that Teq6 may actually be a member of another subgroup of elements which were underrepresented in this study. However the copy number of this subgroup would be much less than the copy number of the major subgroups represented here (Table 4.1).

The copy number of Teq31 was also estimated as shown by Figure 4.7. Teq31 has nucleotide sequence identity of 71% with Teq6 and these two sequences are closely related to each other as far as the sequence identity is concerned. Form Figure 4.8 it can clearly be seen that hybridization signal of the genomic DNA slot 6 is approximately equal to the hybridization signal intensity of the control slot 4 which contains 2.3 $x10^8$ targets or in other words it contains 10^4 more copies of Teq31. However Figure 4.8 panel a shows that the hybridization signal of genomic DNA slot 6 is slightly higher than the control slot 4, which means that the copy number of Teq31 is approximately between 10,000 and 12,000 copies. The intensity of hybridization signal reduces with an increase in the stringency of wash temperature (panel b, c and d Figure 4.7) suggesting that there may be some other Teq31 type elements in the genome of *Agave tequilana* which have been washed away at higher stringencies. It can be concluded from this result there are other Teq31 type elements in the genome of *Agave tequilana* as the signal intensity reduces with the stringent washes showing heterogeneity in Teq31 type elements. The copy number estimates of Teq31 and Teq6 also suggest that these two elements might be members of the same group but low sequence identity (71% nucleotide and 63% peptide) indicates that they can be members of two different but related subgroups of elements which are heterogeneous in nature and underrepresented in the population isolated in this study.

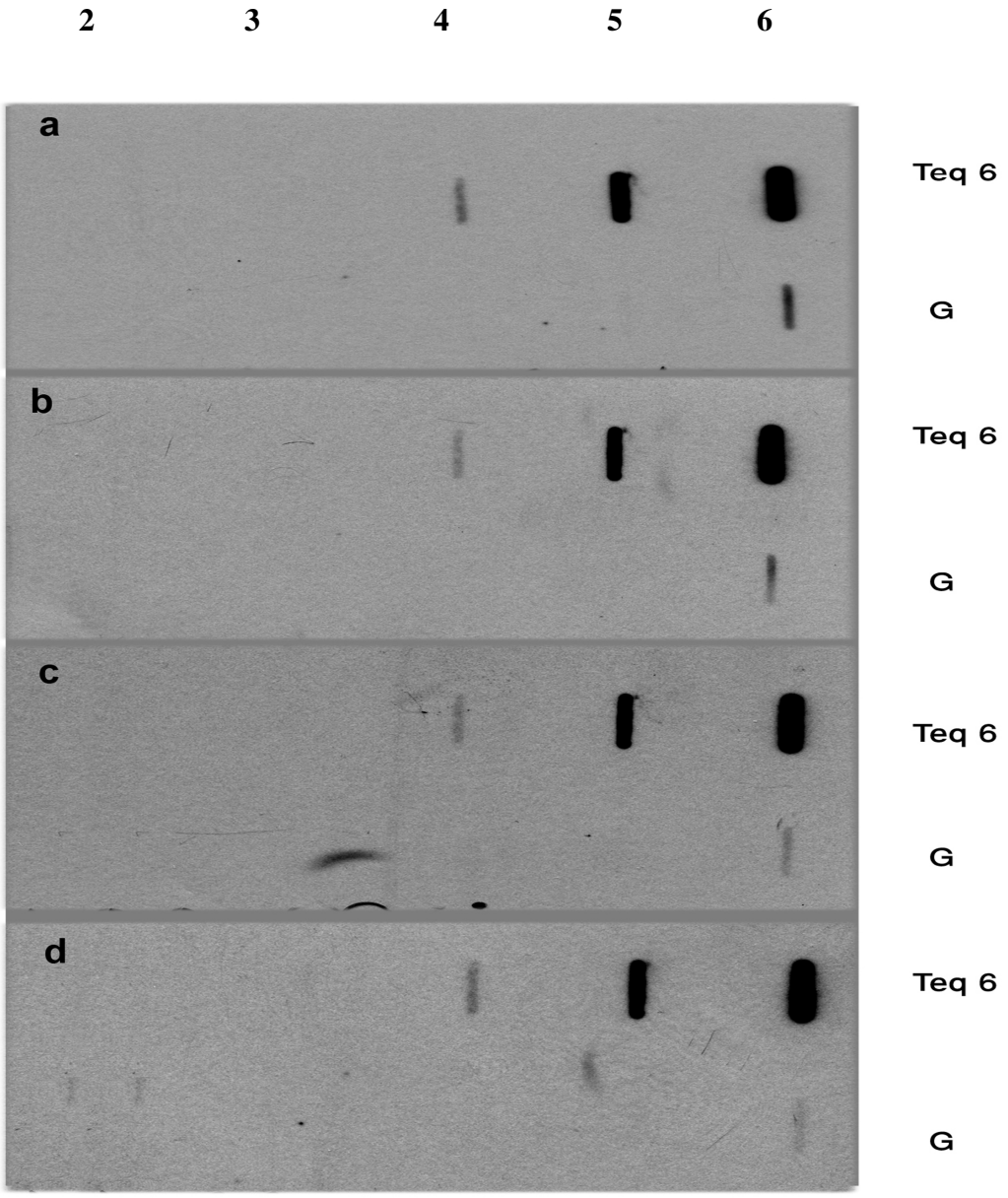


Figure: 4.6 Quantitative slot blot of Teq 6

Teq6 is an ungrouped sequence. Cloned reverse transcriptase fragment derived from Teq6 ($2.3x10^{10}$molecules in slot 6 and $2.3.x10^{9}$ molecules in slot 6 and $2.3.x10^{8}$ molecules in slot 4) were fixed to the membrane. Row G contains genomic DNA from *Agave tequilana (*slot 6 contains 2.3 $x10^{4}$ genomes). Blots were probed and hybridized with cloned reverse transcriptase fragment of Teq 6 and different stringent washes were carried out on the membrane. Panel *"a"* shows the lowest stringency (55°C in 1x SSC), "*b"* and "*c"* represent higher stringencies (60°C in 1x SSC and 60°C in 0.1xSSC) while "*d*" shows the highest stringency (65°C in 0.1xSSC)

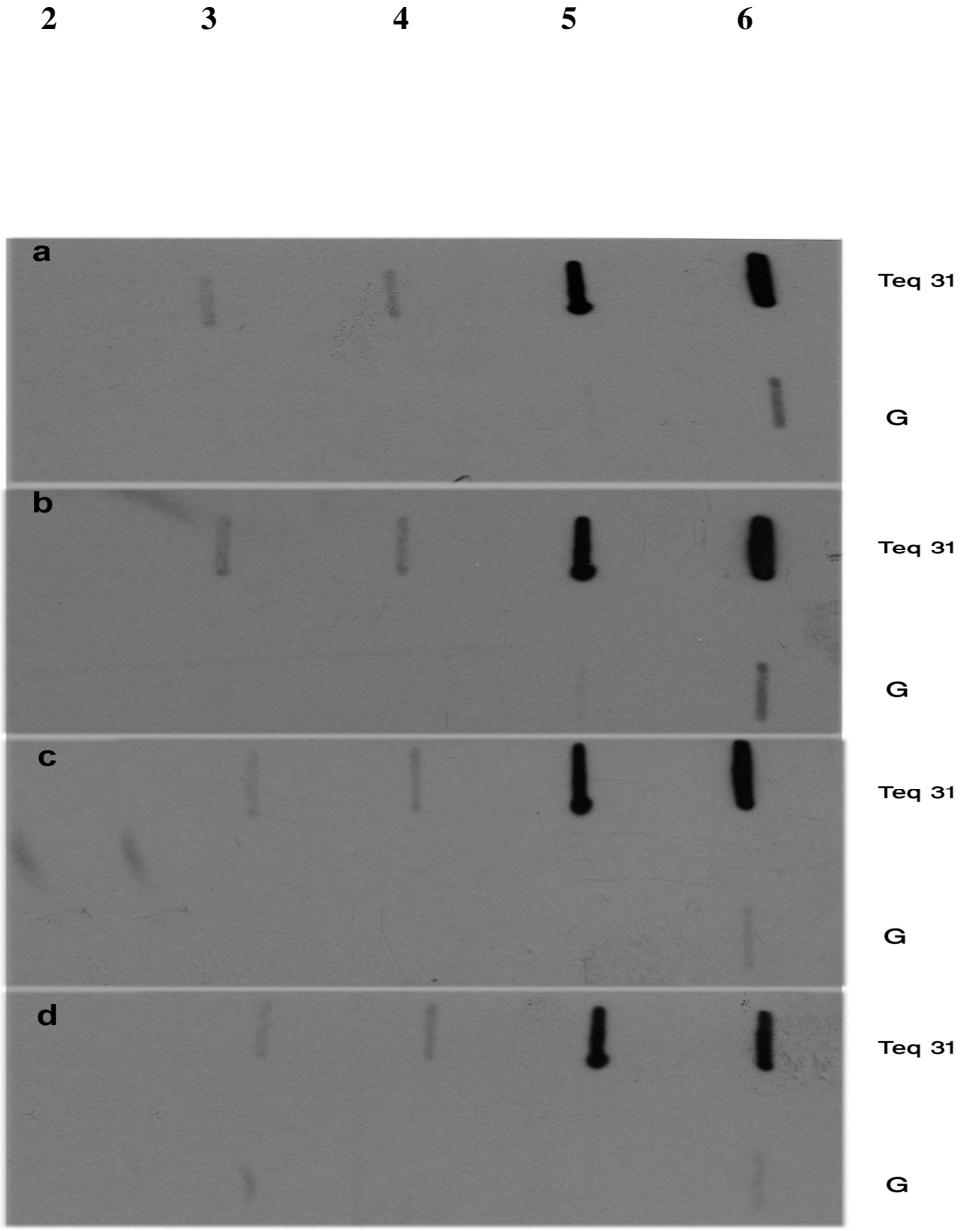


Figure: 4.7 Quantitative slot blot Teq31
Teq 31 is an ungrouped sequence. Cloned reverse transcriptase fragment derived from Teq31 ($2.3x10^{10}$molecules in slot 6 and $2.3.x10^{9}$ molecules in slot 6, $2.3.x10^{8}$ molecules in slot 4 and $2.3.x10^{7}$ molecules in slot 3) were fixed to the membrane. Row G contains genomic DNA from *Agave tequilana (*slot 6 contains 2.3 $x10^{4}$ genomes). Blots were probed and hybridized with cloned reverse transcriptase fragment of Teq31 and different stringent washes were carried out on the membrane. Four panels in the Figureure show different stringent washes performed on the blot. Panel "a" shows the lowest stringency (55°C in 1x SSC), "b" and "c" represent higher stringencies (60°C in 1x SSC and 60°C in 0.1xSSC) while "d" shows the highest stringency (65°C in 0.1xSSC)

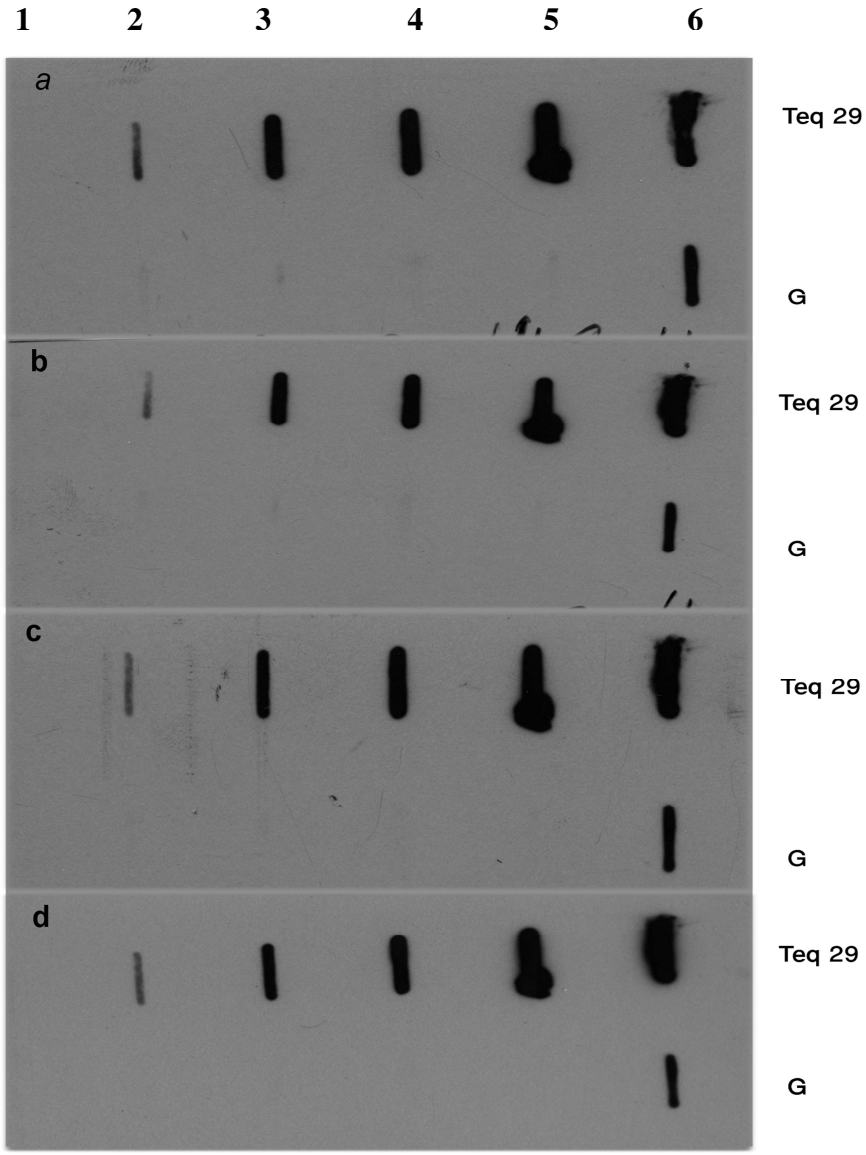


Figure: 4.8. Quantitative slot blot Teq 29
Teq 29 is an ungrouped sequence. Cloned reverse transcriptase fragment derived from Teq29 ($2.3x10^{10}$molecules in slot 6 and $2.3.x10^{9}$ molecules in slot 6, $2.3.x10^{8}$ molecules in slot 4 and $2.3.x10^{7}$ molecules in slot 3) were fixed to the membrane. Row G contains genomic DNA from *Agave tequilana (*slot 6 contains 2.3 $x10^{4}$ genomes). Panels a shows the lowest stringency (55°C in 1x SSC), b and c represent higher stringencies (60°C in 1x SSC and 60°C in 0.1xSSC) while "d" shows the highest stringency (65°C in 0.1xSSC) [previous comments apply]

Teq 29 is another sequence which was considered as an ungrouped sequence in the sequence analysis and phylogenetic analysis. The maximum nucleotide sequence identity that Teq29 has in the present population of sequences is 55% with Teq3. Initially it was thought that Teq29 may belong to Teq3 subgroup as it showed high bootstrap value on the phylogenetic tree as well as some sequence identity with Teq3 subgroup elements. However the sequence identity of Te29 was not more than 55% in case of nucleotide and 54% incase of peptide sequences.Due to these reasons teq29 was also included in the copy number estimates. Figure 4.8 shows the copy number of Teq29 and panels a,b,c, and d on Figure4.8 represent the change in the intensity of hybridization signal at different stringencies. The intensity of hybridization signal of genome slot 6 (row G) containing 2.3 $x10^4$ genomes is approximately equal to (or slightly less than) that of control slot 3 containing 2.3 $x10^7$ molecules, which means that the copy number of Teq29 is around 800 to 1000 copies as shown in Figure: 4.8. There is no significant change in the intensity of the hybridization signal with subsequent stringent washes on Teq29 (panel b, c and d). So Teq29 is an individual low copy number element in the genome of *Agave tequilana* or it belongs to a subgroup of elements which is related to Teq3 subgroup but is homogeneous in nature.

Teq22 has a 64% nucleotide sequence identity with Teq40 and 61% nucleotide identity with Teq14 and Teq15 which are the members of Teq2 subgroup. Teq22 also showed nucleotide sequence identity of 57% with Teq20 and Teq21 but it had a long branch length on the phylogenetic tree. Although Teq22 seems to be related to Teq2 subgroup elements but like other ungrouped elements Teq22 has low nucleotide sequence identity with Teq2 subgroup elements and its nucleotide sequence was even less similar to the rest of the sequences in this population going down to as low as 8% (with Teq18) and (9% with Teq16). Figure 4.9 shows the copy numberestimation of Teq22 as The hybridization signal of 2.3 $x10^4$ genomes (Row G slot 6) is same as the hybridization signal of 2.3 $x10^9$ target molecules of control DNA for Teq22 is approximately equal to or slightly less than hybridization signal of 2.3 $x10^9$ target molecules of control DNA. It was estimated that the copy number of Teq22 is between 60,000 and 80,000 copies in the genome of *Agave tequilana.* With a closer look at the Figure 4.9 it can be seen that the intensity of hybridisation signal is decreased with increasing stringency (panel c and d, Figure 4.9) which means that Teq22 is not a single sequence but there may be more Teq22 type sequences which would be washed away with an increase in the wash temperature. Another possible reason fro a decrease in the signal could be the Tm of the probe but the Tm of Teq22 is 79°C which is well above the maximum wash temperature (65°C). So it can be suggested that Teq22 is a member of high copy number heterogeneous subgroup of elements underrepresented in this study. Like Teq22, Teq41 is another element with longer branch on the phylogenetic tree.

Teq41 shoes a nucleotide sequence identity of 57% with Teq37 and 56% with Teq36 but its sequence similarity with the other elements in the population is low. Figure 4.10 shows the copy number of Teq41 and panels a, b, c, and d show the intensity of hybridization signal at different stringencies. The copy number of Teq41 is approximately 100 000 copies; however the intensity of the signal only slightly changes with an increase in the wash temperature. Nevertheless it decreases sharply at the highest stringency (Figure 4.10 panel d). So Teq41 is also represents a high copy number heterogeneous group of elements which are underrepresented in the population of elements here in this study.

The estimation of retrotransposon copy number and heterogeneity in the population of sequences isolated in this study (chapter3) revealed that Ty1-copia retrotransposons occure as high copy number heterogeneous population of retrotransposons in the genome of *Agave tequilana* as anticipated. It has also confirmed the presence of high copy number subgroups of the sequences but suggested that the ungrouped sequences on the isolation tree may actually be other subgroups of numerous heterogeneous elements instead of single elements shown by the phylogenetic tree.

In another study conducted on the isolation and characterization of RNaseH domain sequences of *Ty1-copia* retrotransposons, clusters of high copy number closely related sequences were found (personal communication with Alaxandros Bousious). The results of this chapter and findings about RNaseH domain sequences lead to the idea of joining RT and RNase H domain sequences by PCR to get a clearer picture of abundance of Ty1-copia retrotransposons in the genome of blue agave.

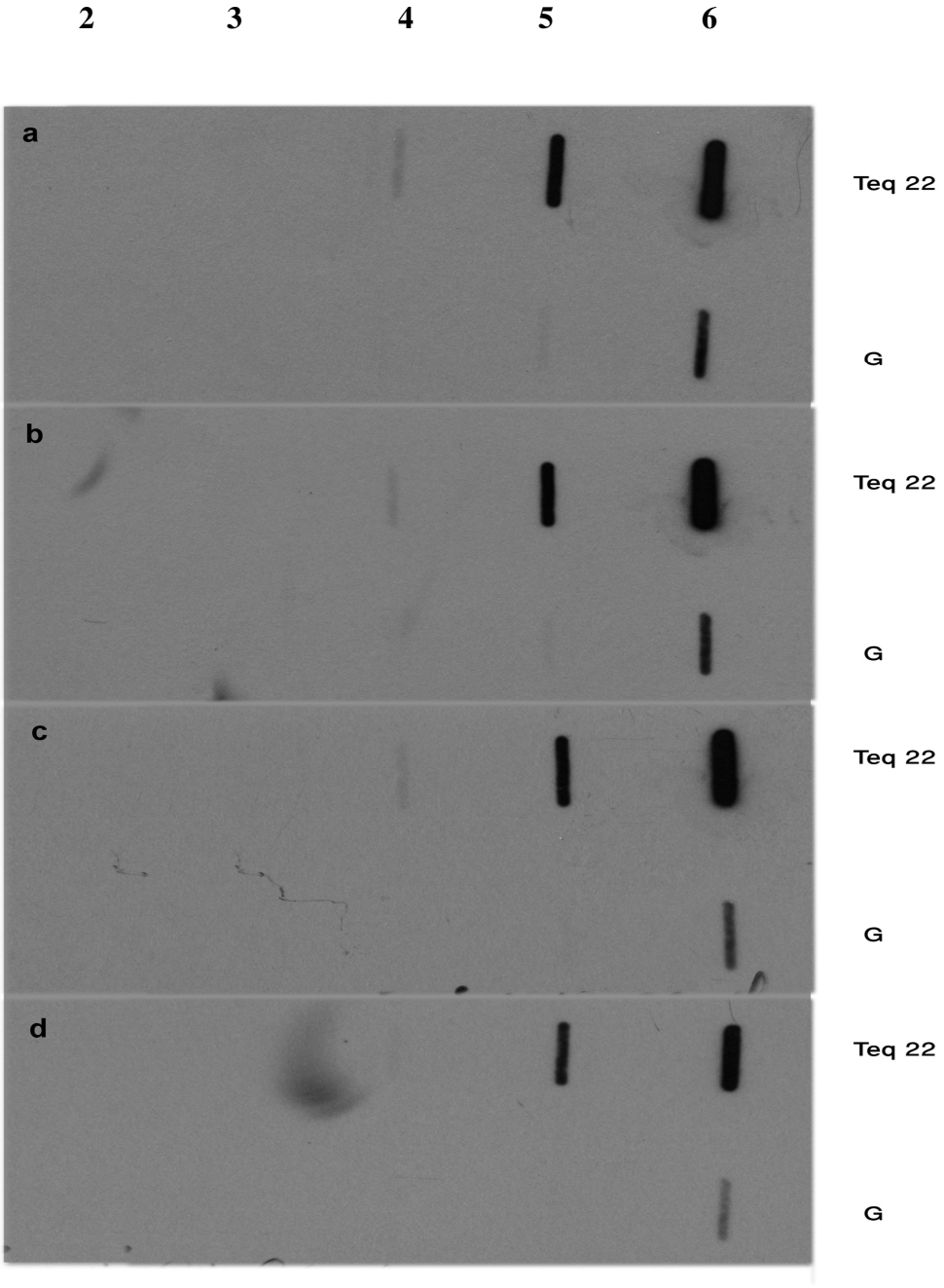


Figure: 4.9 Quantitative slot blot Teq 22

Teq 22 is an ungrouped sequence. Cloned reverse transcriptase fragment derived from Teq22 ($2.3x10^{10}$ molecules in slot 6 and $2.3.x10^{9}$ molecules in slot 6 and $2.3.x10^{8}$ molecules in slot 4) were fixed to the membrane. Row G contains genomic DNA from *Agave tequilana (*slot 6 contains 2.3 $x10^{4}$ genomes). Blots were probed and hybridized with cloned reverse transcriptase fragment of Teq 22 and different stringent washes were carried out on the membrane. Panel "*a"* shows the lowest stringency (55°C in 1x SSC), "*b*" and "*c"* represent higher stringencies (60°C in 1x SSC and 60°C in 0.1xSSC) while *d* shows the highest stringency (65°C in 0.1xSSC)

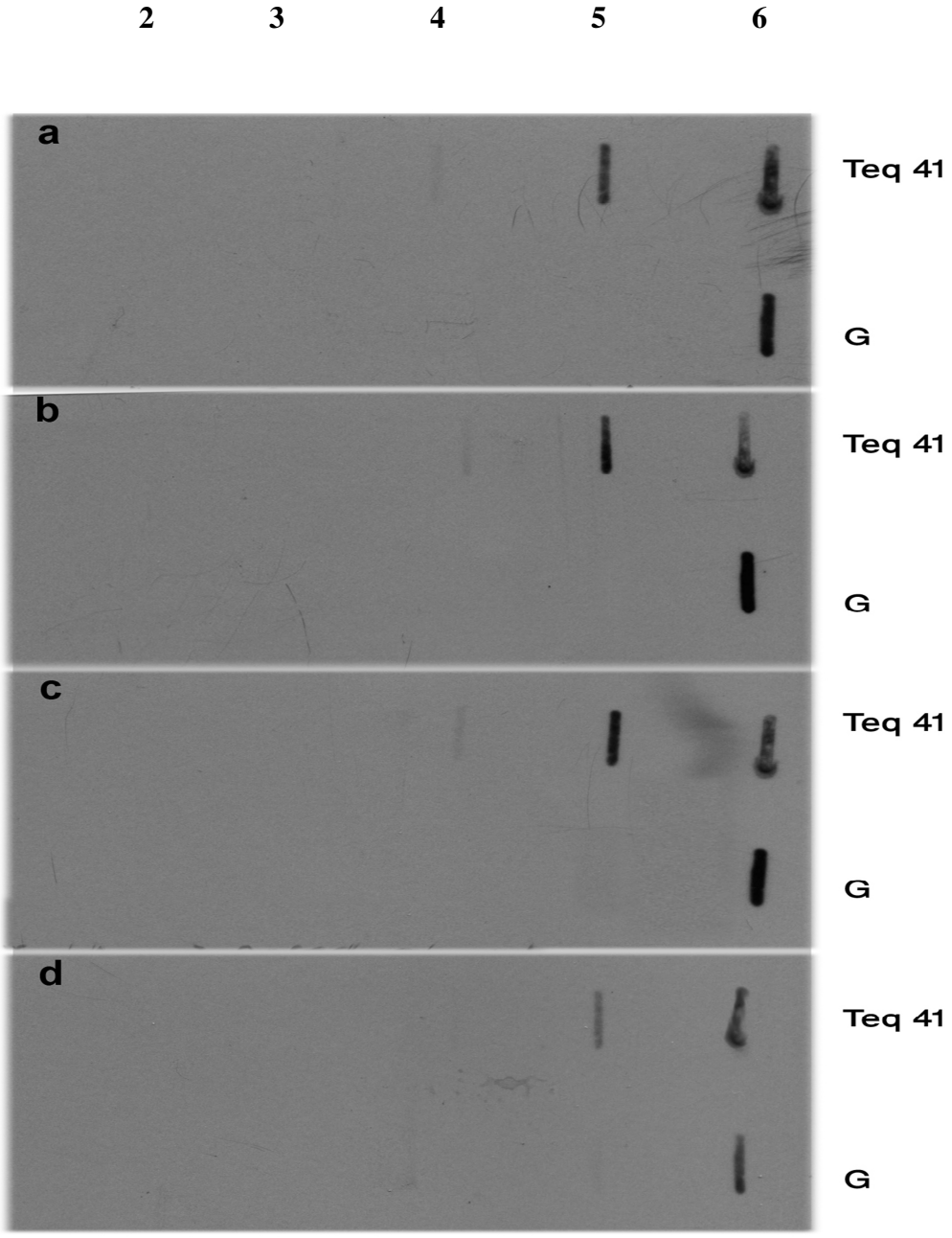


Figure: 4. Quantitative slot blot Teq41

Teq 41 is an ungrouped sequence. Cloned reverse transcriptase fragment derived from Teq41 ($2.3x10^{10}$ molecules in slot 6 and $2.3.x10^{9}$ molecules in slot 6 and $2.3.x10^{8}$ molecules in slot 4) were fixed to the membrane. Row G contains genomic DNA from *Agave tequilana (*slot 6 contains $2.3\ x10^{4}$ genomes). Blots were probed and hybridized with cloned reverse transcriptase fragment of Teq 41 and different stringent washes were carried out on the membrane. Panel "*a"* shows the lowest stringency (55°C in 1x SSC), "*b"* and "*c"* represent higher stringencies (60°C in 1x SSC and 60°C in 0.1xSSC) while "*d*" shows the highest stringency (65°C in 0.1xSSC)

Subgroup	Probes	Estimated copy number
Teq1	Teq1	100,000 copies
Teq2	Teq5	10,000 copies
Teq3	Teq3	100,000 copies
Teq24	Teq33	100,000 copies
Ungrouped sequences	Teq22 Teq41 Teq31 Teq6 Teq29	60,000 - 80,000 copies 100,000 copies 10,000- 12000 copies 10000copies 800 -1000 copies

Table 4.1: Copy number of different subgroups and copy number of ungrouped elements

Name of the probe	Melting temperature (Tm)	Name of the probe	Melting temperature (Tm)
Teq1	77 ℃	Teq29	83℃
Teq3	82 ℃	Teq31	80℃
Teq5	81℃	Teq33	78℃
Teq6	82℃	Teq41	77℃
Teq22	79℃		

Table 4.2: Melting temperature (Tm) of probes.

4.3.7: Joining RT sequences with RNaseH sequences by PCR

In the joining up experiment different subgroups of RT sequences and major subgroups of RNaseH sequences were joined by PCR using a downstream RT primer and an upstream RNaseH primer. In total eight primers were designed from the RNaseH domain sequences and five primers were designed from different subgroups of RT sequences. The primer sequences are given in the general materials and methods (chapter2 section 2.4).

A series of PCR reactions was carried out using all possible combinations of the primers from RT and RNaseH domains. Interestingly only one group of RT sequences joined up with one group of RNaseH sequences (Teq5 subgroup of RT with A21 subgroup of RNaseH). Different combinations of primers used in this experiment are shown by table 4.3.

RT primers	**RNaseH Primers**							
Teq37 RTJ	A8RHJ	A15RHJ	A18RHJ	A22RHJ	A2RHJ	A9-5RHJ	A21RHJ	A1RHJ
Teq30 RTJ	A8RHJ	A15RHJ	A18RHJ	A22RHJ	A2RHJ	A9-5RHJ	A21RHJ	A1RHJ
Teq3 RTJ	A8RHJ	A15RHJ	A18RHJ	A22RHJ	A2RHJ	A9-5RHJ	A21RHJ	A1RHJ
Teq5 RTJ	A8RHJ	A15RHJ	A18RHJ	A22RHJ	A2RHJ	A9-5RHJ	A21RHJ	A1RHJ
Teq21 RTJ	A8RHJ	A15RHJ	A18RHJ	A22RHJ	A2RHJ	A9-5RHJ	A21RHJ	A1RHJ

Table 4.3: Primer combinations used in the joining up PCR

RTJ represents the downstream primers from RT domain while RHJ represents the upstream primers from RNaseH domain of Ty1-copia retrotransposons. The green shade shows the primer combination wich amplified the fragment between the two regions.

In the joining up PCR every primer fro RT domain was used incombination with every primer from the RNaseH domain. So the joining up of RT and RNaseH domain sequences revealed that only Teq5 group elements are similar to A21 sequences from the RNaseH domain. It also suggested that the genome of *Agave tequilana* contains more elements than estimated in this chapter and the retrotransposons population of blue agave is moch diverse and abundant than presented in this chapter.

4.4 Discussion

In this chapter a detailed investigation of *Ty1-copia* retrotransposon copy number has been carried out. The reverse transcriptase (RT) sequences of the retrotransposons used as probes in a slot blot analysis have previously been isolated and characterized in the chapter 3. A detailed copy number estimation of different subgroups revealed that all of the major subgroups are highly abundant and heterogeneous in the genome of *Agave tequilana.* On the other hand the copy number of ungrouped elements estimated in this chapter suggests that the ungrouped elements might not be single elements, they could be subgroups of other elements related to them which were underrepresented in the population of sequences isolated in this study. Some of the single elements like Teq22, Teq41 and Teq31 occur in high copy number (100,000 copies of Teq41 and 60,000-80,000 copies of Teq22) but they do not cluster together with any of the major subgroups. The high copy number and heterogeneity exhibited in slot blot analysis of these individual sequences suggests that they might belong to other subgroups which could not be isolated in this piece of work.

Three of the four major subgroups (Teq1, Teq24 and Teq3) contain around 100,000 copies while the population of Teq2 subgroup is comparatively low with 1000- 10000 copies in the genome of *Agave tequilana*. The individual Ty1-*copia* retrotransposons that do not belong to the major subgroups, are also in high copy number with Teq22 and Teq41 estimated to contain approximately about 60,000-80,000 and 100,000 copies respectively. Teq31 contain around 10,000 - 12000 copies each and the population of Teq6 is around 10000 copies. Teq29 is a low copy number individual element with an approximate copy number of 800-1000 copies.

The average length of Ty1-*copia* group retrotransposons is about 5.5kb (Vitte and Panaud, 2005), if all of these elements are considered to be full length elements than the percentage proportion of *Agave tequilana* genome occupied by Ty1-*copia* elements can be calculated on the basis of its genome size (2C= 8.8 pg or 8624 Mbp). On this basis the Teq1, Teq3 and Teq24 subgroups with approximately 100,000 copies would each occupy about 6.37% of the *Agave tequilana* genome (~19% combined) while Teq2 subgroup being less numerous occupies around 0.63% of the genome. Teq41 is another high copy number element which can possibly occupy another 6.37% of the genome. Teq22 is an element with 60,000 to 80,000 copies occupying about 3.8% - 5.1% while Teq31 can occupy around 0.63% - 0.76% and Teq6 can occupy 0.63% of the *Agave tequilana* genome. Teq29 is a low copy number element occupying around 0.063% of the genome. In total four major groups of the elements in this study can occupy up to 19.74% and all

of the elements included in this study can make up to 32% of the genome of *Agave tequilana*. This estimate is based on the hypothesis that all of the retrotransposons included in this study are intact and full length elements. However it is now well established that retrotransposons could be removed from the genome by different removal mechanisms such as unequal homologous recombination and illigtimate recombinations (Bennetzen et al., 2005). The intra- element (between two LTRs of the same element) and inter-element (LTRs of two different but homologous elements) homologous recombination results in the formation of solo LTRs. The generation of solo LTRs is a major mechanism for the removal of LTR retrotransposons (Bennetzen et al., 2005). The unequal homologous recombination and resulted solo LTR formation is partially responsible for the deletions from the *Arabidopsis* (Devos et al., 2002). On the other hand high relative ratio of solo LTR to intact elements in rice (Ma et al., 2004) suggests that the process of solo LTR formation through unequal homologous recombination may be more active in rice than *Arabidopsis* (Bennetzen et al., 2005). The repetitive DNA has also been removed permanently through illegitimate recombination (Vicient et al. 1999 ;(Devos et al., 2002; Pereira, 2004) causing a complete reshuffling of intergenic regions (San Migual et al. 2002 ; Wicker et al 2003a). In barley unequal homologous recombination of both types (inter-element recombination and intra- element recombination) have occurred but inter-element recombination is considered to be the mechanism responsible for the contraction of massive retrotransposon amplification (Shirasu et al., 2000; Vitte and Panaud, 2005). In case of *Arabidopsis* solo LTR formation through unequal homologous recombination as well as formation of deletions through illegitimate recombination appear to have contributed to reduce the genome size (Devos et al., 2002; Vitte and Panaud, 2005). Illegitimate recombination has previously been recommended as a counter balancing force against amplification of retrotransposons and reason for genome size differences in *Drosophilla* (Petrove and Harti, 1998; Petrove et al. 2000). However the mechanism of unequal homologous recombination and solo LTR formation has been considered to be too weak to counteract the massive amplification of LTR retrotransposons in maize (Bennetzen and Kellogg, 1997; SanMiguel et al., 1998). Moreover there are many nonautonomous retrotransposons that completely or partly rely on the proteins expressed by other elements elsewhere in the genome (Vitte and Panaud, 2005). Non autonomous retrotransposons with defective coding regions or no coding capacity are found in all classes of the transposable elements(Bureau and Wesler 1994;Wicker et al.2003b;(Kalendar et al., 2004). Additionally many copies of the elements with defective coding regions with frameshifts or in frame stop codons may be the offspring of functional elements. This can be due to the error prone nature of the reverse transcription or accumulated mutations after the insertion

in to the genome (Keulen et al. 1997; (Gabriel et al., 1996). The percentage of agave genome occupied by Ty1-*copia* elements might significantly be reduced by an active deletion mechanism like unequal homologous recombination in *Agave tequilana.* However the copy number estimates here in this chapter revealed that all of the Ty1-copia retrotransposns have not been amplified. Because copy number and heterogeneity estimates of the ungrouped sequences here suggested that they may not be single ungrouped elements but members of other subgroups of elements which were not fully represented in the population of elements in this study. The results of joining up of RT and RNaseH domain described here also confirmed that the population of elements discussed here does not represent the actual status of the *Ty1-copia* retrotransposons in the genome of *Agave tequilana*. So the copy number estimated here might actually be an underrepresentation of the actual copy number of Ty1-copia retrotransposons in *Agave tequilana.*

Although counter balancing mechanisms like homologous recombination and illegitimate recombination occure in the genomes eukaryotes but different mechanisms are found to be active indifferent organisms for example unequal homologous recombination is responsible for the solo LTR formation and deletion of retrotransposons in the genome of barley but it is not common in the maize genome. On the other hand illegitimate recombination is considered to be responsible for the complete reshuffling of retrotransposons in *Arabidopsis* (Devos et al., 2002) , wheat (Wicker et al., 2001) and *Drosophila* (Petrove and Hartl, 1998 ; Petrove et al., 2000). Nevertheless there must be some mechanism of counter balancing to keep the genome functional. As agave has a relatively large genome and despite the function of the various retrotransposon deletion mechanisms, it is more probable that a large fraction of agave genome is composed of LTR retrotransposons, similar to the proportion in the large genomes of maize, 50-80% (Meyers et al., 2001; Sanmiguel and Bennetzen, 1998) and barley 70% (Feschotte et al., 2002; Vicient et al., 1999b). Moreover the error prone nature of the reverse transcription can give rise to a population of closely related heterogeneous retrotransposons, such heterogeneous populations of closely related elements have been characterised in a variety of plants (Matsuoka and Tsunewaki, 1999; Pearce et al., 1996b; Price et al., 2002; Stuart-Rogers and Flavell, 2001; Voytas et al., 1992). Same sort of heterogeneous population of closely related elements clustering together was observed here (Figure 4.1). The slot blot analysis of *Ty1- copia* retrotransposons revealed that only four major subgroups of elements can account for one fourth of the genome of agave. However this does not show the real picture of the genome because there may be other retrotransposons that are yet to be isolated and characterized. In conclusion we can say that a large proportion of the genome of *Agave tequilana* is composed of *Ty1-copia* retrotransposons.

CHAPTER 5

Retrotransposon expression and somaclonal variation in *Agave tequilana*

5.1: Introduction

Retrotransposons transpose through the reverse transcription of an RNA intermediate. They are ubiquitous constituents of eukaryotic genomes and have played a significant role in the structure, organization and evolution of plant genomes (Bennetzen, 2000; Kumar and Bennetzen, 1999; Wessler et al., 1995). Due to their replicative mode of transposition retrotransposons can successfully amplify and accumulate to high numbers in plants, which can often lead to a significant expansion of the host plant's genome size (Bennetzen and Kellogg, 1997) .

Most retrotransposons are thought to be transcriptionally inactive (Kumar and Bennetzen, 1999) or silent in somatic tissues but active during certain stages of plant development and under the effect of stressful conditions (Grandbastien, 1998) for example *Tnt1* is expressed in roots at very low level (Pouteau et al., 1991) while activity of *Prem2* elements of maize has been detected only in early microspores (Todorovska, 2007).

5.1.1: Somaclonal variation and retrotransposons

In vitro clonal propagation of plants provides a method to efficiently multiply and maintain large number of elite plant genotypes and is extensively used in the regeneration of genetically modified plants ((Mhiri et al., ; Mhiri et al., 1997). Genotypic instability is commonly observed in plants derived from tissue culture and is thought to be induced by the tissue culture procedure (Evans et al., 1984; Larkin and Scowcroft 1981). This phenomenon of genotypic variation in the clones of plants is termed somaclonal variation whilst mostly introducing undesirable effects (Larkin and Scowcroft 1981) can be used in breeding to increase genetic variability (Karp 1995). The molecular basis of somaclonal variation is not precisely known, but both genetic and epigenetic mechanisms have been proposed to be responsible (Smykal et al., 2007). Somaclonal variations may arise as a result of point mutation, rearrangement of nuclear or organellar DNA, polyploidy, the activation of mobile elements (transposable elements) or epigenetic changes causing deviation from the desired phenotype quality standard (Phillips et al 1994; Jaligot et al 2000). Although the molecular basis of somaclonal variation is largely unknown a commonly accepted hypothesis is the breakdown of normal cellular controls resulting in the genetic and

epigenetic instabilities, hence alterations in the gene expression (Phillips et al. 1994; Kaeppler et al. 2000). One of the possible explanations of somaclonal variation relates to the dynamic structure of the genome, activities of highly abundant and temporary mobile DNA elements such as transposons and retrotransposons ((Kazazian, 2004; Kidwell and Lisch, 1997). Many retrotransposons are activated by stress and environmental factors and appear to be a major and broad source of genetic variation from chromosomal alteration to the tuning of gene expression (Kazazian, 2004; Kidwell and Lisch, 1997). It is important to note that transposable elements not only cause mutations by insertions or excision at a single chromosomal site but also cause changes at the genomic level associated with the transposition (Kidwell and Lisch, 1997). The genetic variability resulting from the transposable elements ranges from changes in the whole genome to changes in the individual nucleotides which may produce major effects on the phenotypic traits or small changes detectable only at the DNA sequence level (Britten 1996). Transposable elements can cause changes by inserting into exons of the host genes (Kidwell and Lisch, 1997) for example *P* elements of Drosophila (Rubin et al., 1982) and the Ac-Ds family of maize (Wessler et al., 1987). They can also insert into introns and heterochromatic regions as well as regulatory regions of genes (Kidwell and Lisch, 1997). Insertion is not however the only way retrotransposons can cause mutations they can also create variation by mediating recombination which can result in structural changes of chromosomes (Banga et al., 1991). It is well established that transposable elements are a source of genetic variation (Kidwell and Lisch, 1997) however over millions of years of evolution, they have achieved a balance between detrimental effects on the individuals and long term beneficial effects on a species through genome modification (Kazazian, 2004). The activity of retrotransposons can be induced by stresses and a consequence of increased retrotransposon mobility is the creation of new genetic variability that can be useful in stressful conditions (Capy et al., 2000).

5.1.2: Stress activation of retrotransposons in plants.

Many of the plant retrotransposons are activated by various biotic and a abiotic stresses (Grandbastien, 1998). *Tnt1* is the most studied retrotransposon with respect to transcriptional activity as well as stress activation. Initially it was suggested that the expression of *Tnt1* is confined to specific conditions such as freshly isolated protoplasts and roots of tobacco (Pouteau et al., 1991). It was later reported that transcription of Tnt1 can also be greatly stimulated by different microbial elicitors of plant defence responses for example factors of fungal origion

(crude extracts of *Trichoderma viride a*nd elicitins purified from *phytophthora* species) and culture supliments of bacterium *Erwinia chrysanthemi* (Pouteau et al., 1994). The transcription of *Tnt1* is not limited to tobacco; it can also be induced in other plant species for example tomato and *Arabidopsis* by wounding and various other abiotic stress factors (Mhiri et al., 1997). Mechanical injuries such as are induced during the punching out of leaf disks during tissue culture can induce expression of *Tnt1* in *Arabidopsis* and tomato however the expression is reported to be strong but highly localized (Mhiri et al., 1997). On the other hand the*Tnt1A* promoter is activated by ozone air pollution in tomato but not in tobacco (Pourtau et al., 2003). Transcriptional activity of *Tnt1* has been shown to be regulated and strictly controlled (Casacuberta and Santiago, 2003).

The promoter of Tnt1A contains two different boxes, located in the U3 region of the LTR, that have been shown to be important for the element's transcription and show sequence similarities with plant defence promoters (Vernhettes et al., 1997). Tnt1B and Tnt1C differ from Tnt1 in their U3 region that probably controls their expression.

Tto1 and Tto2 are two other elements of tobacco reported to be activated by stresses such as tissue culture (Hirochika, 1993). Apart from the detection of Tto1 and Tto2 an increase in the copy number of these element was also reported due to the activation under tissue culture conditions (Hirochika, 1993). Moreover Tto1 has been reported to be autonomously capable of transcription (Hirochika et al., 1996b) . As *Tnt1, Tto1* is also activated by wounding and methyl jasmonate (Takeda et al., 1998). In the case of Tto1, a13 bp motif has been identified as a cis regulatory sequence associated with the induction of Tto1 expression in defence related stresses (Takeda et al., 1999). *Tos17* is a retrotransposon of rice which is not only activated by tissue culture but also causes mutations in tissue culture cells (Hirochika et al., 1996a) and is regulated at the transcriptional level in rice (Hirochika, 1997). It was also found that Tnt1 of tobacco is capable of transposition and it can transpose into different loci when introduced into *Arabidopsis* (Lucas et al., 1995). On the other hand *Tto1* of tobacco was found to be actively transposing in the cultured cells of rice (Hirochika et al., 1996b) which indicates that the factors responsible for the transcriptional activity of *Tnt1* and *Tto1* are conserved in monocots as well as dicots (Hirochika et al., 1996b). Until recently *Tnt1*, *Tto1* and *Tos17* were thought to be the only retrotransposons demonstrating transcriptional and transpositional activities in plants (Hirochika et al., 1996a; Hirochika et al., 1996b).

Transcriptional activity has also been reported in other plant retrotransposons, for example *CIRE1*, *Reme1* and *TLC1.1* are three newly identified LTR retrotransposons active in citrus,

melon and tomato respectively. The expression of *CIRE1* is induced by wounding as well as after the exogenous application of methyl jasmonate and auxin (Rico-Cabanas and Martinez-Izquierdo, 2007) the transcription of Reme1 is induced by UV light (Ramallo et al., 2008) while *TLC1.1* is activated in response to multiple stresses (Salazar et al., 2007). Additionally in the activationof BARE1 under tissue culture conditions has also been reported to be the cause of genetic instability in wild barley (Li et al., 2007).

5.1.3: The role of retrotransposons in somaclonal variation and the generation of genetic variation.

Many investigations have been made to investigate the effect of biotic and abiotic stresses on transcriptional as well as transpositional activity of retrotransposons in a variety of organisms including plants. Different stress agents have been used to induce retrotransposon activity in plants such as those involved in protoplast isolation and tissue culture (Hirochika, 1993; Hirochika et al., 1996a; Rico-Cabanas and Martinez-Izquierdo, 2007), UV light (Ramallo et al., 2008), wounding (Mhiri et al., 1997), microbial elicitors and fungal pathogens (Mhiri et al., 1999; Pouteau et al., 1994) and methyl jasmonate (Takeda et al., 1998). The post stress genetic and epigenetic changes in retrotransposons have been detected by a variety of molecular biology tools. Stress induced transcription has commonly been investigated by RT-PCR (Ramallo et al., 2008; Rico-Cabanas and Martinez-Izquierdo, 2007). However some researcher have studied genetic variability caused by the activation of retrotransposons by using SSAP (sequence specific amplification polymorphism) and IRAP (inter retrotransposon amplification polymorphism) (Grandbastien et al., 2005; Vicient et al., 2001). Estimation of copy number after stress activation has also been used to measure the impact of stress on retrotransposons (Hirochika, 1993).

In vegitatively propagated plants, meristems are undifferentiated cell lineages, derived from mitosis. For plants with clonal propagation, meristem gives rise to vegitatively produced offshoots and mitotically derived decendants. These plants do not undergo meiosis so the only source of new genetic variation has to be somatic mutations; therefore vegitatively propagated plants should have low levels of genetic variation (Callaghan et al., 1992). However differences in different cultivars of olive, which is usually propagated by asexual means, have been demonstrated (Mekuria et al., 1999). Clonal diversity has also been reported in a herbaceous plant *Bryonia alba* (Novak et al., 2000) and a clonal grass *Clamagrostis porteri* (Esselman et al., 1999).

Agave is native to Mexico (Gentry, 1982) Due to its long life cycle the blue agave has been propagated through asexual means for the last 200 years (Valenzuela-Zapata, 1997). Initially it was reported that there is no genetic variability in the cultivated blue agave (Vega et al., 2001), however this report triggered research for the evaluation of genetic diversity in other agave species and asexual genetic variability was detected in *Agave fourcroydes* by AFLP (Infante et al., 2003). Asexual genetic variability was also reported in different agave species using AFLP as well as ISTR (inverse sequence tag repeat) and it was suggested that the genetic variation was created during the clonal propagation of these plants (Infante et al., 2006). In another attempt of evaluating asexual genetic variability, it was proposed that the gene pool of *Agave tequilana* is not so narrow as was previously thought (Vega et al., 2006).

Retrotransposon-based molecular marker techniques such as SSAP and IRAP (Inter Retrotransposon Amplification Polymorphisms) show the insertion profiles of particular retroelements. The work of this chapter uses these methods along with mRNA analysis to investigate the insertion profiles and activation of retrotransposons in agave tissue culture lines and during the vegetative propagation of agave through the study of mother and daughter plants.

5.2: Results

One of the aims of this chapter was to evaluate the asexual genetic variability among different parent agave plants and their clonally propagated baby plants. Evaluation of genetic variability among different tissue culture lines due to stress activation of retrotransposons was the other main aim of thic chapter. The evaluation was carried out by using retrotransposon based molecular markers like IRAP and SSAP. The detailed methodology is described in 2.8 of chapter 2

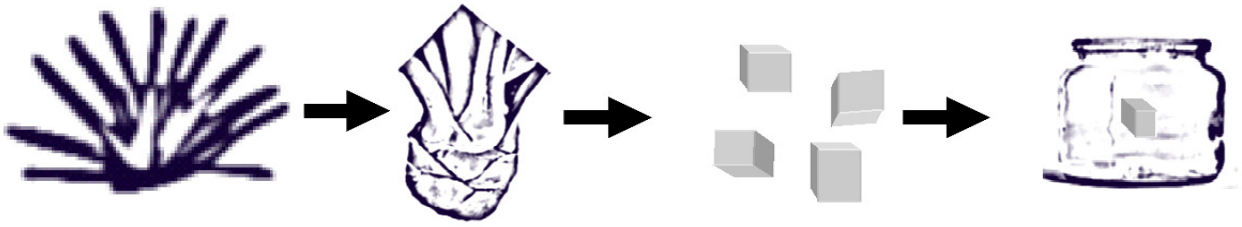
5.2.1: Activity of Ty1-*copia* retrotransposons in *Agave tequilana.*

The investigation was carried out by the assessment of transposon activity in parent and daughter plants using SSAP and IRAP (details in general material and methods chapter 2). SSAP relies on the amplification of DNA between a retrotransposon integration site and nearby restriction site with a ligated adapter (Waugh et al., 1997). In contrast IRAP does not rely on restriction sites and amplifies the DNA between two LTR sequences using a single outwards facing retrotransposon primer, (Kalendar et al., 1999). Retrotransposons tend to insert into non genic regions The

densities of transposable elements can be so high that often elements are inserted within other elements giving a highly "nested" structure (Bennetzen, 2000). When retrotransposons occur in close proximity, frequency of recombinations is increased which results in deletions.

The study material for this section was a population of *A. tequilana* mother and daughter plants which were already being cultivated at the University of Sussex. Previous work (Ivan Saldana) involved a population of agave plants which had derived from a single Mexican plant. In the course of the previous study these had grown to a size where they were producing their own daughter plants. This population of mother plants and daughter plants were considered to be a useful model population for the study of retrotransposon activity. As many retroelements have been reported to be active under tissue culture conditions a number of tissue culture lines were also set up from *A. tequilana* plants growing at Sussex.

To investigate genetic instabilities caused by stress in agave a tissue culture was established using explants from the central head or "Pinna" of agave.



agave plant

central head or Pinna of the agave plant

explants from the merismatic tissue of central head

cultured explants

The newly produced tissue culture accessions were used as a starting material to study stress induced changes in agave. Firstly genomic DNA from different tissue culture accessions was isolated and retrotransposon insertional polymorphisms among these tissue culture accessions estimated by IRAP. Secondly total RNA was isolated from the tissue culture to perform northern blotting by using retrotransposon RT (reverse transcriptase) sequences isolated in the previous chapter as probes. Finally transcriptional activity was studied by conducting RT-PCR experiment with degenerate RT primers. To carry out RT- PCR mRNA was isolated from Total RNA; this mRNA was reverse transcribed into cDNA which was then used as a template in RT- PCR. Apart from insertional polymorphism studies and RT-PCR a normal AFLP was also carried out to see if the variations among tissue culture accessions were general or retrotransposon specific.

Collectively speaking different molecular biology techniques were combined to investigate the retrotransposon expression, asexual genetic diversity and transcriptional activity in *Agave tequilana.*

5.2.2: Insertional polymorphism of Ty1-copia element "A1" in vegitatively propagated *Agave tequilana* plants.

Two LTR specific primers have been used; one of these primers belongs to the high copy number A1 subgroup of retrotransposons while the second primer belongs to the A17 subgroup of elements (Bousious et al., 2007).Figure 5.1 shows the insertion profiles of the A1 retroelement (Bousious et al 2007) using IRAP. In Figure5.1 the new IRAP band are shown by black arrows and lost bands or deletions are shown by red arrows. It is clear that there is variation among different parent and baby agave plants. 1b, 2b and 3b, baby plants show at l one, two and three deletion respectively as IRAP bands have been deleted in these baby plants. However 3b shows an extra IRAP band which could mean a retroelement has inserted in baby 3 as compared to its parent plant. Baby 4b shows one new insertion as well as a deletion compared to parent 4 while four new insertions can be seen in baby 5b. New insertions are also present in 6b1 and 6b2 as well as 7b1 and 7b2. Interestingly four IRAP band are deleted in baby plant 8b. Parent 7 also contains fewer bands compared to the other parent plants but its baby plants show three new insertions in the form of three new IRAP bands.

5.2.3: Insertional polymorphism of Ty1-copia element "A17" in vegitatively propagated *Agave tequilana* plants.

.

.The genetic variability with respect to retrotransposons was also evaluated by IRAP with a primer from A17 family of Ty1-*copia* retrotransposon (Bousious et al., 2007). Figure 5.2 Shows retrotransposon polymorphism of agave retroelement A17 in parent and baby agaves. It is clear that A17 is less polymorphic in agave than A1 family of retrotransposons. This family of retrotransposon was initially reported to be less polymorphic and less abundant in agave than A1 family of Ty1-copia retrotranspsons (Bousious et al., 2007). All of the parent and baby plants share common bands along with some deletion as well as new insertions, for example baby 2b and 4b contain 1 deletion each while 4b and 8b contain some new insertions. However the presence of common IRAP band in unrelated plants can also be seen here.The IRAP analysis of parent and baby agave plants revealed retroelement polymorphisms among parent and baby

plants. Figure 5.1 and Figure 5.2 clearly showed losses of original bands which are shown by red arrows in the Figures. Some baby plants showed new insertions or they regained the original bands that they might have lost in a previous generation of vegetative reproduction.

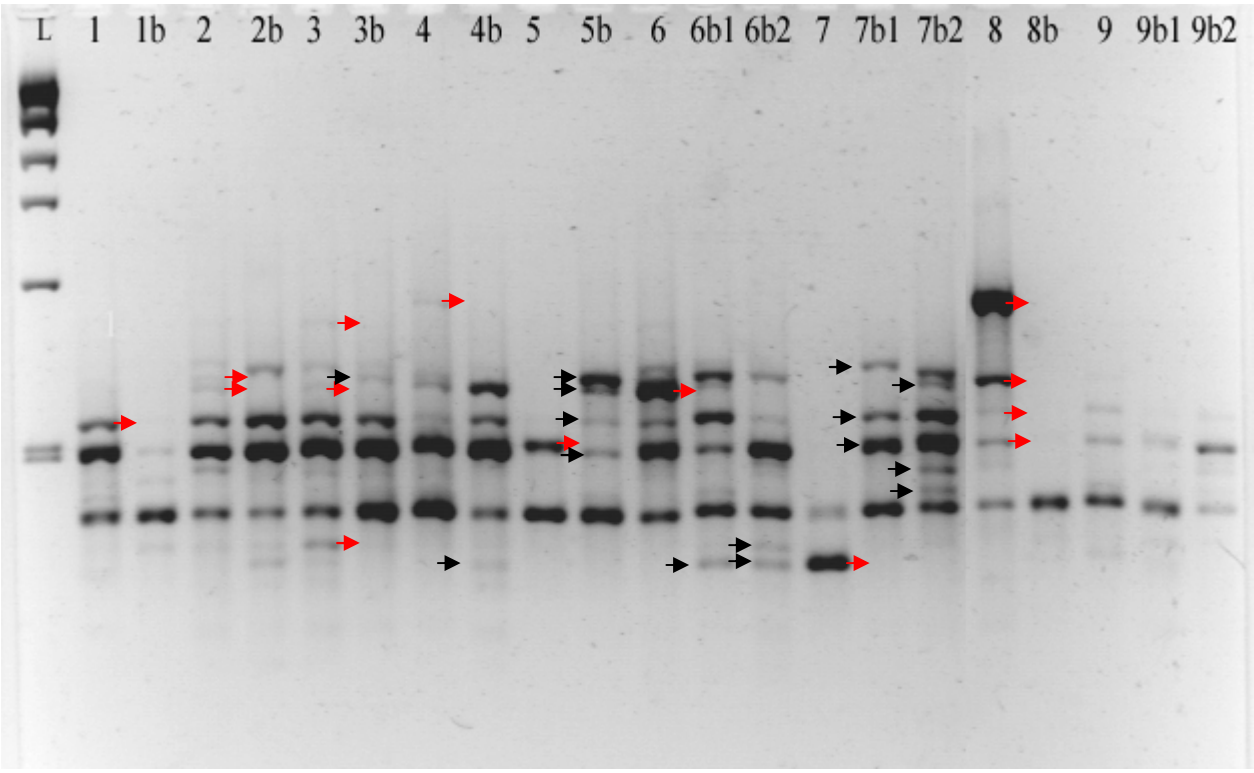


Figure: 5.1 IRAP analysis of retroelement A1 polymorphism in parent and baby agave plants.

The IRAP analysis was carried out by PCR using retrotransposon A1 LTR specific primer. The parents are represented by numbers. Daughter (baby) plants are represented by a number with "b" while L = 100bp ladder (NEB). Missing bands are indicated by a red arrow and each new band is represented by a black arrow.

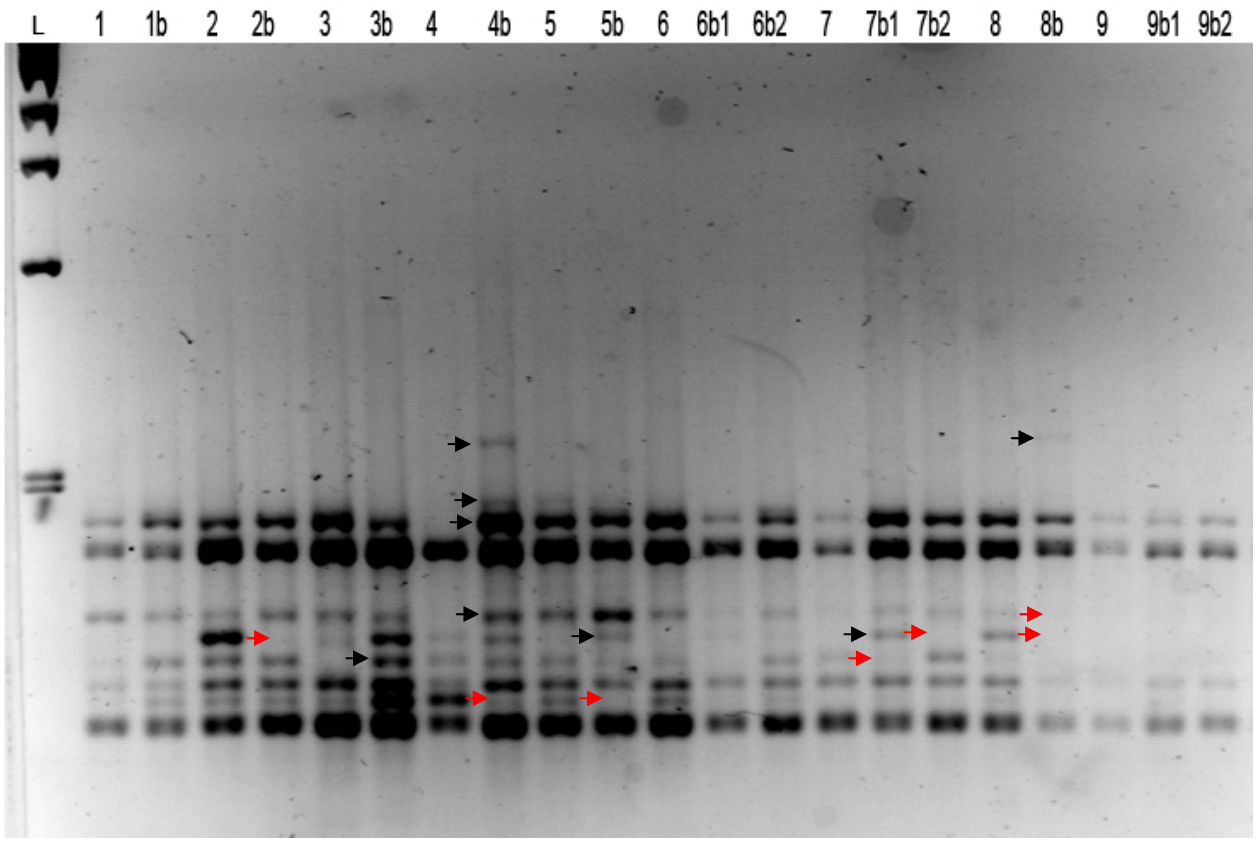


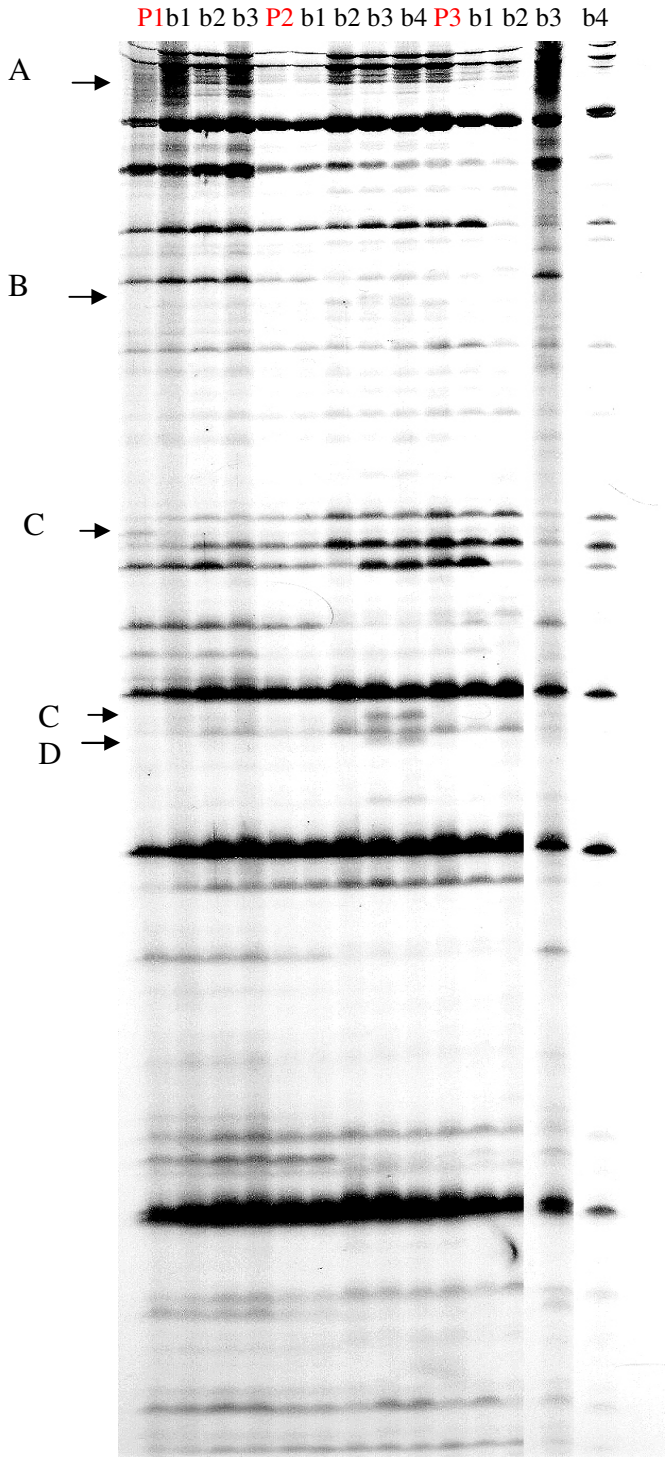
Figure: 5.2 IRAP analysis of A17 polymorphism in parent and baby agave plants.
IRAP analysis of parent and baby agave plants reproduced by vegetative propagation was carried out by PCR using LTR specific primer A17. The parents are represented by numbers while their baby plants are represented by a number with "b" while L represents the 100bp ladder. Each lost IRAP band between parent and baby plants is represented by a red arrow and each new band is represented by a black arrow.

5.2.4 SSAP analysis of parent and baby *agave tequilana* plants

Parent and baby plants of *Agave tequilana* were also compared by SSAP analysis using a combination of retrotransposon specific primer and Eco + GC while the preamplification was carried out by Mse +CG and Eco + G. SSAP was first used to see the distribution and location of BARE1 retrotransposon in the barley genome (Waugh et al., 1997). In principle this is a modification of standard AFLP (Vos et al., 1995), however the final amplification is performed between retrotransposon integration site and a restriction site to which an adapter has been ligated (Waugh et al., 1997). SSAP has been used to study retrotransposon activation and mobility, biodiversity and genome evolution, mapping of genes and estimation of genetic distances and assessment of essential derivation of varieties (Schulman 2007). However the requirement of digestion of DNA and ligation of adapters as well as use of radioactively labeled primers and sequencing gels makes SSAP complicated and labour intensive. In comparison IRAP is a simple method which does not require any digestion or ligation of adapters and the PCR products can be separated on an agarose gel (Kalendar and Schulman 2006). In the work described here SSAP was initially used to evaluate genetic variability in parent and baby plants but it was eventually replaced by IRAP.

The SSAP was carried out using LTR specific primer from the high copy number subgroup A1 (Bousious et al., 2007). Figure 5.3 shows the pattern of SSAP bands in parent and baby agave plants. Retrotransposon polymorphism can be seen in the form of new bands in the baby plants or loss of bands from the baby plants. Figure 5.3 confirmed the presence of new insertions in the baby plant b3 and b4 of the parent plant P2 (Figure5.3) while band B was deleted from the baby plant b1 of parent P1. These extra bands confirm the presence of new insertions in baby 7b1 and 7b2 (Figure 5.1). An interesting finding from SSAP analysis is that the new insertions are at the same point in both of the baby plants.

Collectively speaking evaluation of genetic variability among parent and their vegitatively propagated baby plants revealed that losses as well as gain of retrotransposons. Our results show that baby plants not only contain new insertions of retrotransposons but also contain deletions .In some cases for example baby plant 8b a loss of four IRAP bans has been shown (Figure 5.1 and 5.2) such a phenomenon has never been reported in any other plant species. In conclusion retrotransposon-based molecular markers show that asexual genetic diversity does exist in vegitatively propagated agave plants as reported before (Infante et al., 2003; Infante et al., 2006). However the origin of this genetic variability and the reasons behind it are yet to be found.



Fiugure5.3: SSAP analysis of retroelement A1 insertions in vegetatively propagated agave plants. .
Three parent plants were compared to their baby plants. The parents are shown as P1, P2 and P3 while their baby plants are shown as b1, b2, b3 and b4. The polymorphic bands are represented by an arrow with a letter(A,B,C).The SSAP analysis was carried out by using a combination of a retrotransposon specific primer (A1) and *Eco*+GC while the preamplification was done by *Mse*CG + *Eco*G.

5.2.5: Tissue culture mediated activation of retrotransposons and retrotransposon polymorphism in agave.

This section deals with the retrotransposon polymorphism in agave tissue culture. The existence of somaclonal variations in different tissue culture accessions of agave and their relationship with stress activation of retrotransposon is evaluated.

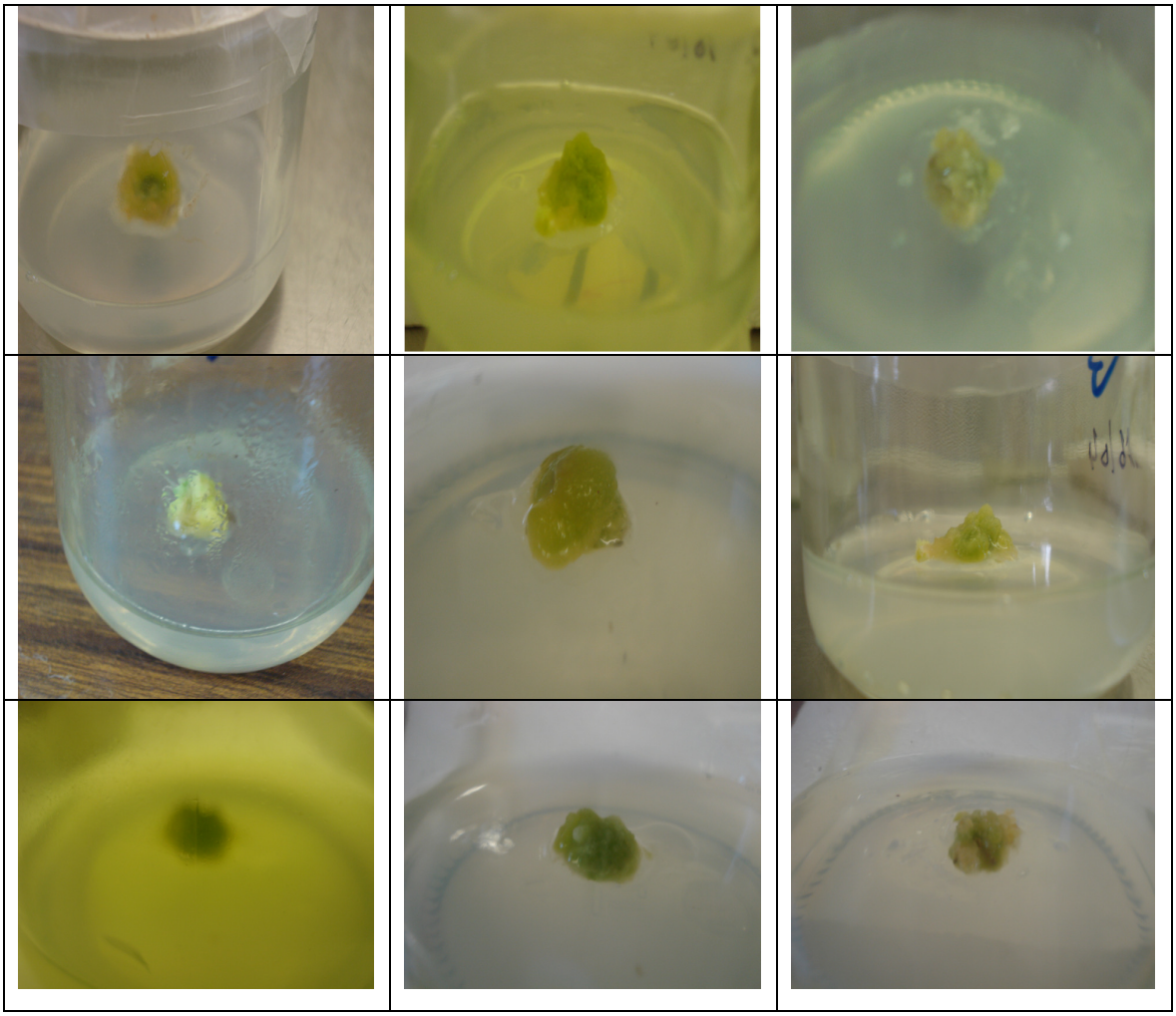
LTR retrotransposons are an important source of genetic diversity and have had a major impact on the structure of plant genomes due to their replicative mode of transposition (Kumar and Bennetzen, 1999). A common feature of retrotransposons is that they are activated by stress and environmental factors and most well characterized retrotransposons are particularly affected by protoplast isolation or *in vitro* cell and tissue culture (Grandbastien, 1998). On the other hand tissue culture induced phenotypic and genotypic variations in plants are collectively called somaclonal variations (Larkin and Scowcroft 1981) and seem to be ubiquitous in plants (Bajaj 1990). It was suggested that the activation of retrotransposons might be responsible for somaclonal variations (Larkin and Scowcroft 1981). Most tissue culture induced mutations are stable and my be explained by the activation of retrotransposons (Hirochika et al., 1996a). Many studies have focused on the activation of retrotransposons in tissue culture for example the activation of *Tto1* and *Tto2* retrotransposons of tobacco (Hirochika, 1993) and *Tos17* retrotransposons of rice (Hirochika et al., 1996a). Over the years a wide variety of molecular biology approaches have been applied to study the activation of retrotransposons and their impact on palnt genomes. Most of the studies were focused on the evaluation of transcriptional activity of retrotransposons which is the first step in the transposition of retrotransposons, while effects of retrotransposon activation like genetic variability and increase in the copy number have also been evaluated. For example the activity of *BARE*1 in grass species has been evaluated by IRAP (Vicient et al., 2001) On the other hand the tissue culture induced activation of *Tos17* in rice was evaluated by northern blotting (Hirochika et al., 1996a). In recent years RT- PCR has extensively been used to see the stress induced transcriptional activity of retrotransposons for example transcriptional activity of *CIRE1* in citrus (Rico-Cabanas and Martinez-Izquierdo, 2007) and *Rem1* in melon has been evaluated by using RT- PCR.

On the basis of work on other systems there is a strong possibility that some elements of the diverse populations of retroelements in agave are activated during stresses associated with tissue culture. Working on the hypothesis that his is the case in agave the retroelement insertional

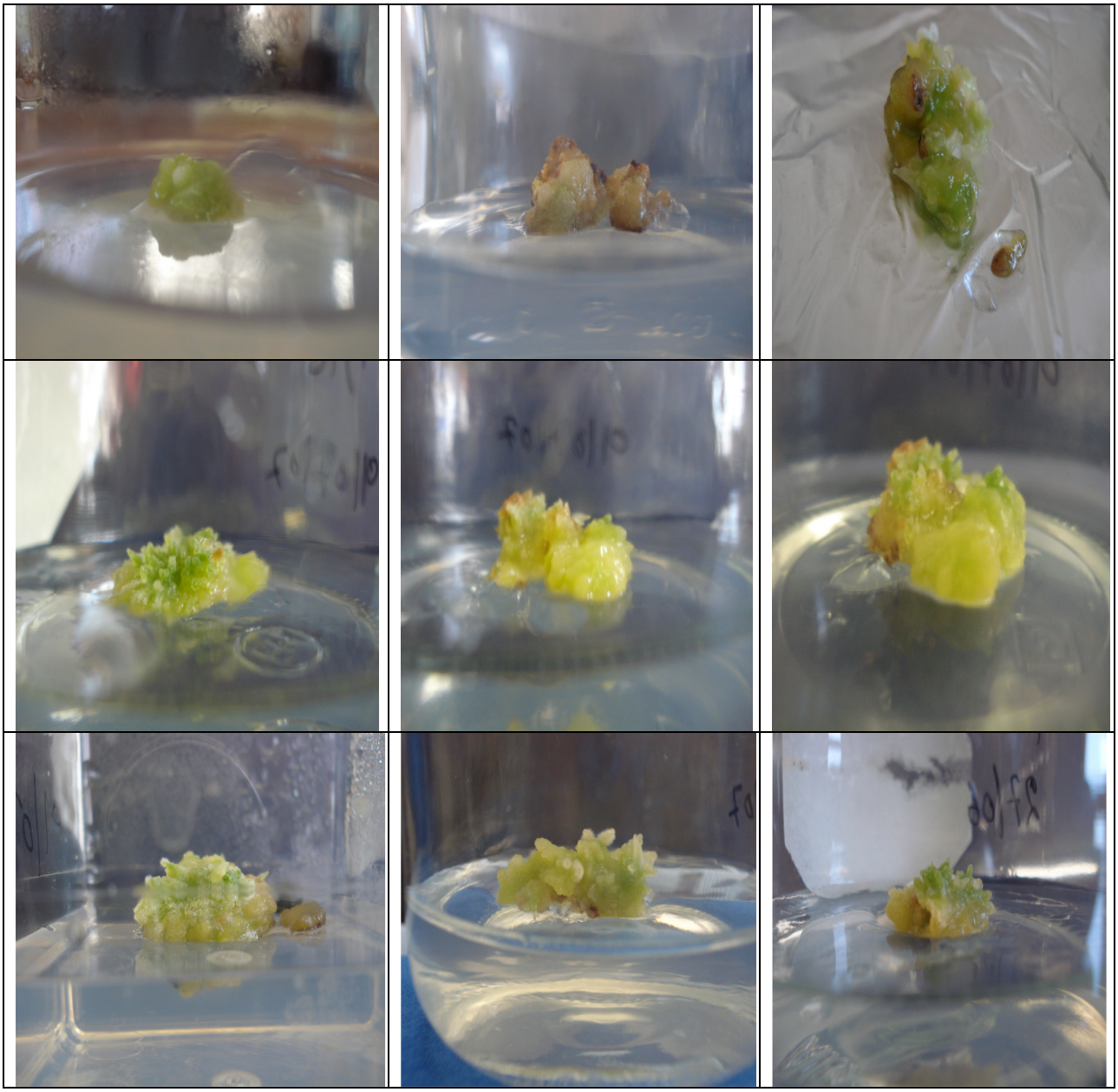
polymorphism in agave tissue culture lines was investigated using retrotransposon-based molecular markers whilst transcriptional activation was investigated by RT-PCR.

5.2.6: Establishment of agave tissue culture lines.

In order to establish a tissue culture of agave, tissues from the central head of an agave plant were used as explants to make use of the meristimatic tissue of centeral head of plant which would help in the induction of callus. Tissue culture media with different concentrations of growth hormones were used to induce callus in agave (for details of tissue culture media, chapter 2, 2.6.5) . Figure 5.4 shows the induction of callus in agave. Although explants from the leaves were also used but the callus was only induced in the explants from central head or "pinna" of agave. In total twenty types of callus induction media were used but callus was induced on only two of them. The callus shown in Figure 5.4 was induced on a medium containing 4.4 micomoles of 2, 4-D, 5.7 μM NAA and 2.2 micomoles of BAP. Figure 5.4B shows the callus growing on a medium containing 22 μM of 2, 4-D, 5.7 μM of NAA, and 4.4 μM of BAP. The callus was growing very fast and it was getting harder as well so the amount of BAP was reduced to 0.88 μM and the temperature of the culture was reduced to 22°C (Figure 5.5) and theses conditions were maintained until the extraction of DNA and RNA (Figure 5.6).



Figureure 5.4 Induction of callus in *Agave tequilana*
Induction of callus by using central head (Pinna) as explant. The medium contained 4.4 micomoles of 2,4-D, 5.7 µM NAA and 2.2 micomoles ofBAP .The culture was grown at 22 °C with a humidity of 50% and 16 hours of light and 8 hours of dark. The cultures were subcultured every three weeks.



Figureure 5.4 B Growth and maintenance of callus
The callus was maintained on a medium with 22 µM of 2,4-D, 5.7 µM of NAA, and 4.4 µM of BAP. The temperature of the growth chamber was 24°C with 50% humidity and 16/8 hours light and dark.

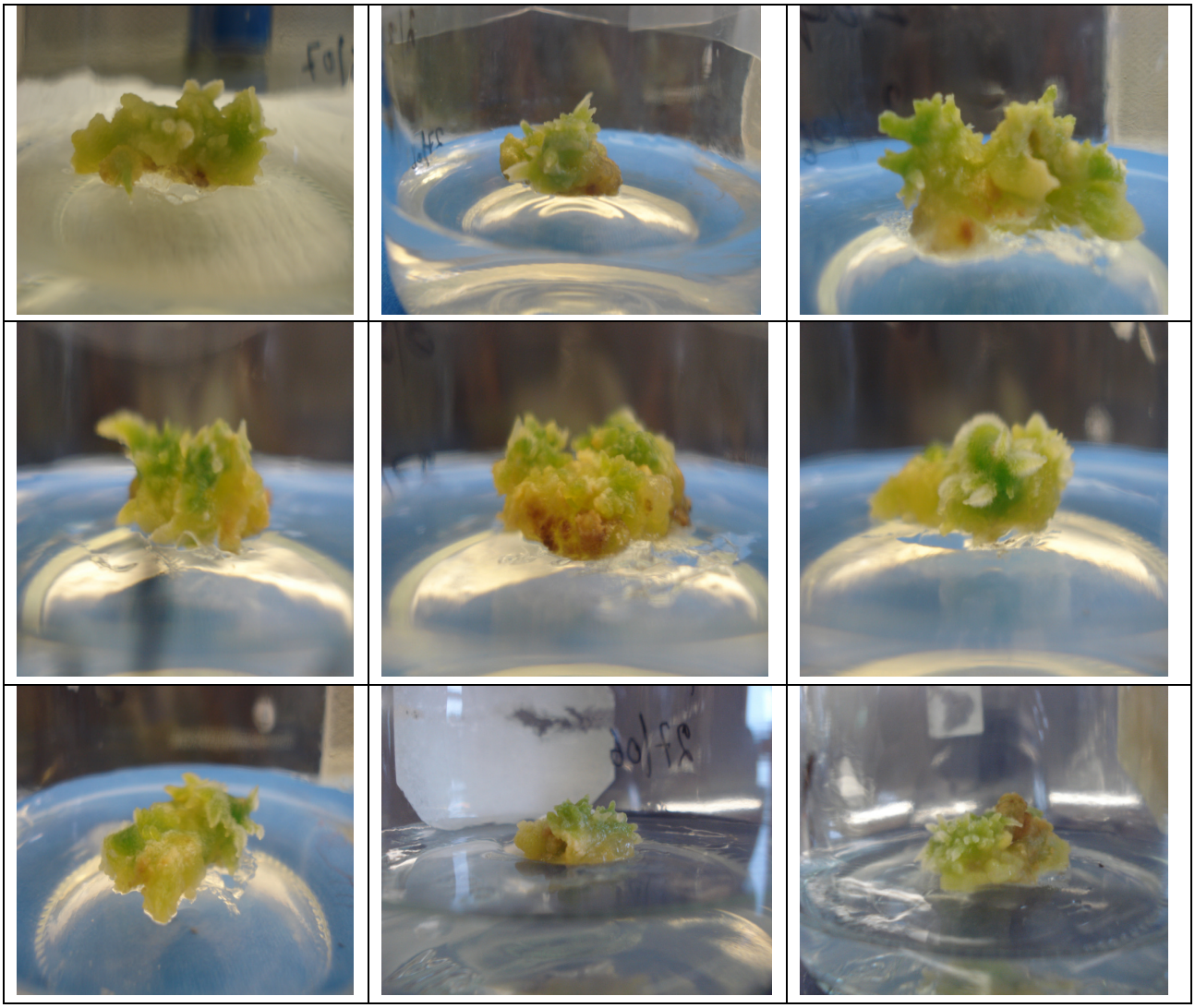


Figure 5.5: Maintenance of callus
The callus was grown on a different medium containing 22.6 µM of 2,4-D, 5.7µM of NAA and 0.88 µM of BAP. The growing conditions were 22°C, 50% humididty and 16 hours/8 hours of light and dark.

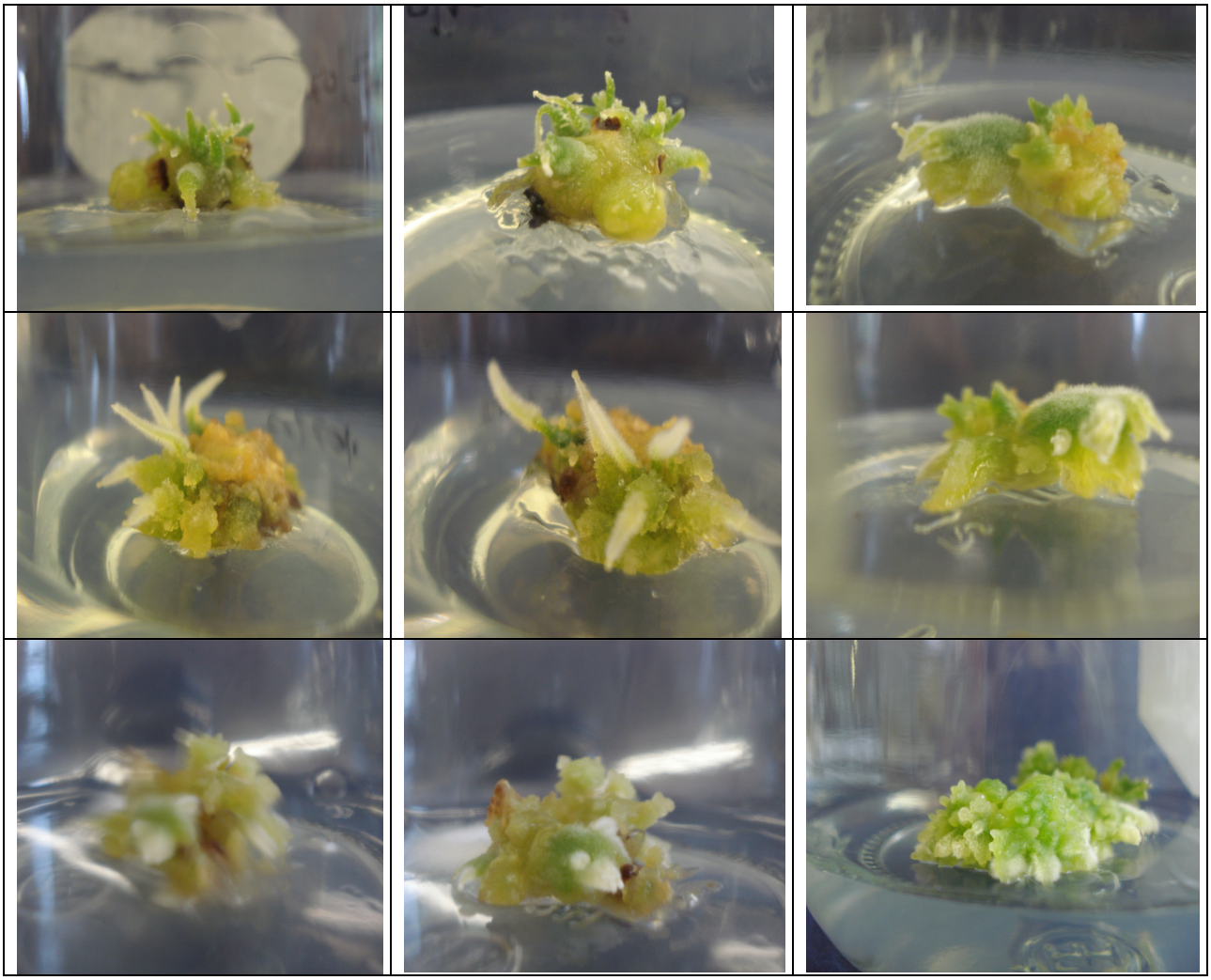


Figure 5.6: The tissue culture used for the extraction of DNA and RNA.
14 weeks old cultures used for the extraction of DNA and total RNA.
Culture was maintained on the medium containing 22 µM of 2,4-D , 5.7 µM of NAA and 0.88 µM of BAP at 22 °C a 50% humidity. The DNA and total RNA was extracted form the 14 weeks old cultures.

5.2.7: Somaclonal variation, and stress activation of retrotransposons in *Agave tequilana*

The ability of retrotransposons to insert into different locations in the genome through reverse transcription of their RNA intermediate make them ideal to be used as molecular markers for the detection of somaclonal variation in plants. Retrotransposons have been used to evaluate the genetic instabilities and somaclonal variations in wild barley (Li et al., 2007) and genetic stability in long term in vitro shoot culture in peas (Smykal et al., 2007). To evaluate the retrotransposon polymorphism caused by the activities of retrotransposons under the effect of tissue culture, IRAP (inter-retrotransposon amplification polymorphism) was used. Figure 5.7 shows the genetic instability due to the effect retrotransposon deletions as well as new insertions. Lane 21 in Figure 5.7 represents the parent plant while lanes 1-20 contain the samples from different tissue culture accessions. In this experiment the IRAP bands were not scored but they were visually scrutinised for any new insertions and possible deletions. With a closer look at Figure 5.7 retrotransposon deletions (lane 1, 4, 18, 19, 20) as well as new insertions (lane 1, 2, 4, 15, 16, 18, 20) can be seen.

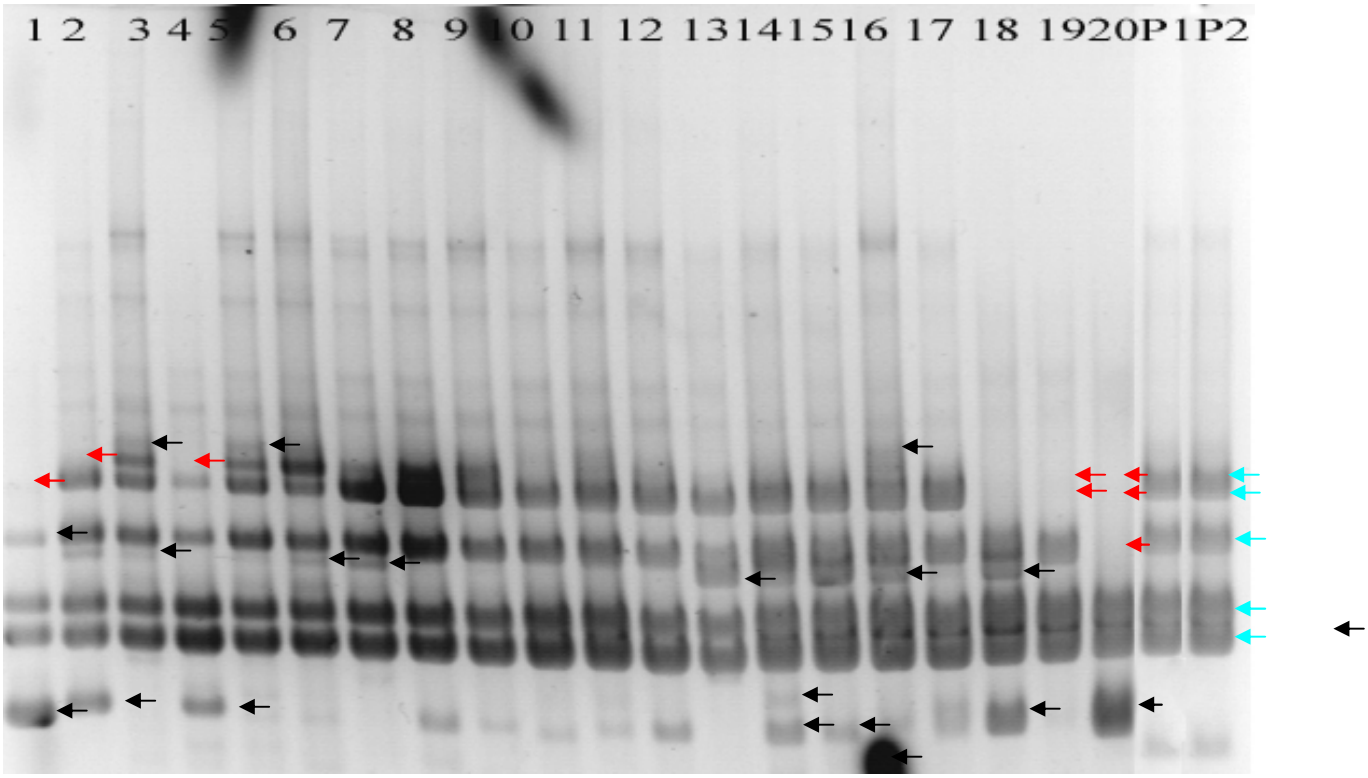


Figure 5.7 Retroelement activity in tissue culture lines using IRAP with A1A primer
Negative of agarose gel. Lanes 1-20 are tissue culture accessions while P1and P2 are two samples from the parent plant. Green arrows represent the IRAP band in the parent plants, black arrows show new insertions while red arrows show deletions.

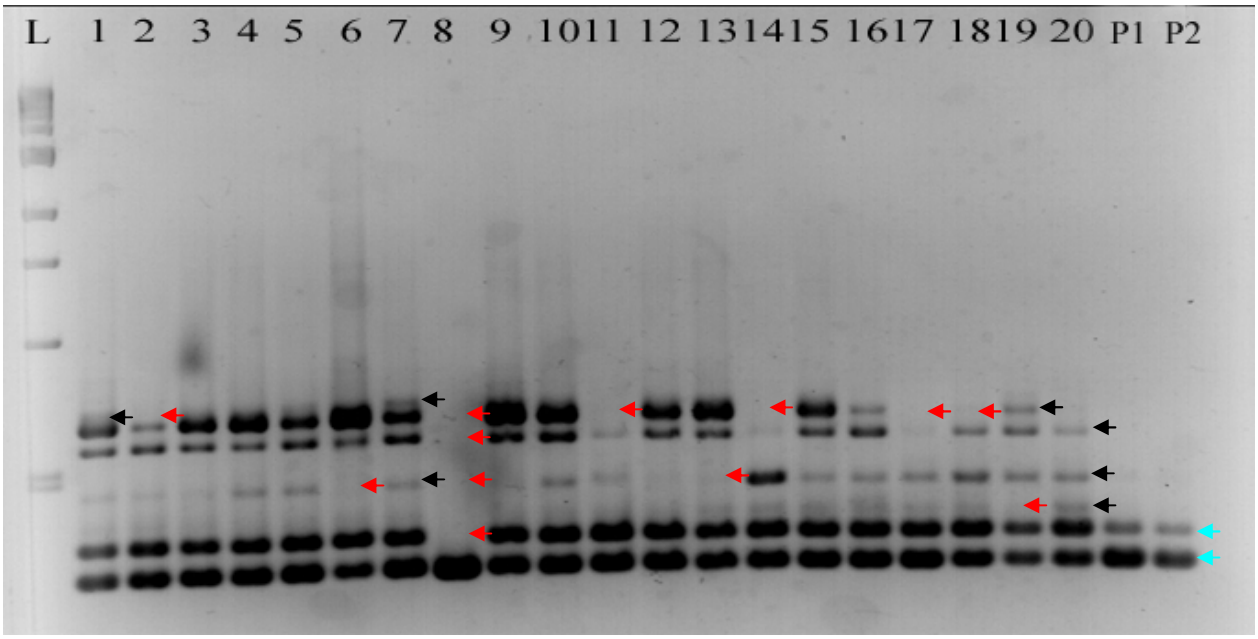


Figure 5.8 Comparative analyses of tissue culture accessions by IRAP with A1B primer

The IRAP analysis was carried out by PCR usinf A1B primer from A1 group of elements.The Figureure is a negative of agarose gel visualized on UV trans-illuminator. Green arrows represent the banding pattern of parent plant samples while each black arrow is a new insertion after tissue culture. Red arrows indicate the deletions compared to the parent plant and among tissue culture accessions.

Figure 5.8 also indicates the presence of new insertions as well as deletion of original copies of retrotransposons (lane 8). From Figure 5.8 it can be seen that the parent plant shows less bands or in other words less copies of retrotransposons as compared to the number of bands in tissue culture accessions. It is evident that a normal vegetative reproduction can cause genetic rearrangement in agave (Figure 5.1, 5.2 and 5.3). As agave has been propagated through vegetative means for last 200 years, the parent plant in this experiment might have lost some of its retrotransposons in the last vegetative propagation event and tissue culture might have induced the transposition of retrotransposons as all of the tissue culture accessions show new bands and share some common bands.

Genetic variability can also be seen from Figure 5.9 which represents IRAP polymorphism with respect to the A17 IRAP primer (2.11 Chapter 2).

Figure 5.9 shows that retrotransposon deletions appear to be more common although there are some new retroelement insertions. Lanes 21 and 22 shows the IRAP banding pattern of the parent plant (green arrows in the Figure) while lanes 1- 20 represent the IRAP banding pattern of tissue culture lines. It can clearly be seen that A17 is also polymorphic in agave tissue culture.

There could be many factors and processes responsible for somoclonal variations and retrotransposons are one of the factors that could be the reason for somoclonal variation. The work described here shows retrotransposons are polymorphic in the different cell lines. It is possible however that the variation may be caused by changes in the DNA sequence surrounding the retroelements. In order to compare the level of polymorphism which was independent of retroelements an, AFLP experiment was carried out on the tissue culture lines. A high level of AFLP polymorphism would indicate that the polymorphism of the retrotransposon-based markers was a symptomatic of other more general genomic changes. Figure 5.10 shows the AFLP banding pattern of parent agave plant compared to the tissue culture lines. AFLP is a robust genetic fingerprinting technique which is based on restriction site polymorphism (Vos et al., 1995). There is no AFLP polymorphism among the tissue culture lines which indicates that that the high levels of polymorphism revealed in the IRAP analysis are retrotransposon specific

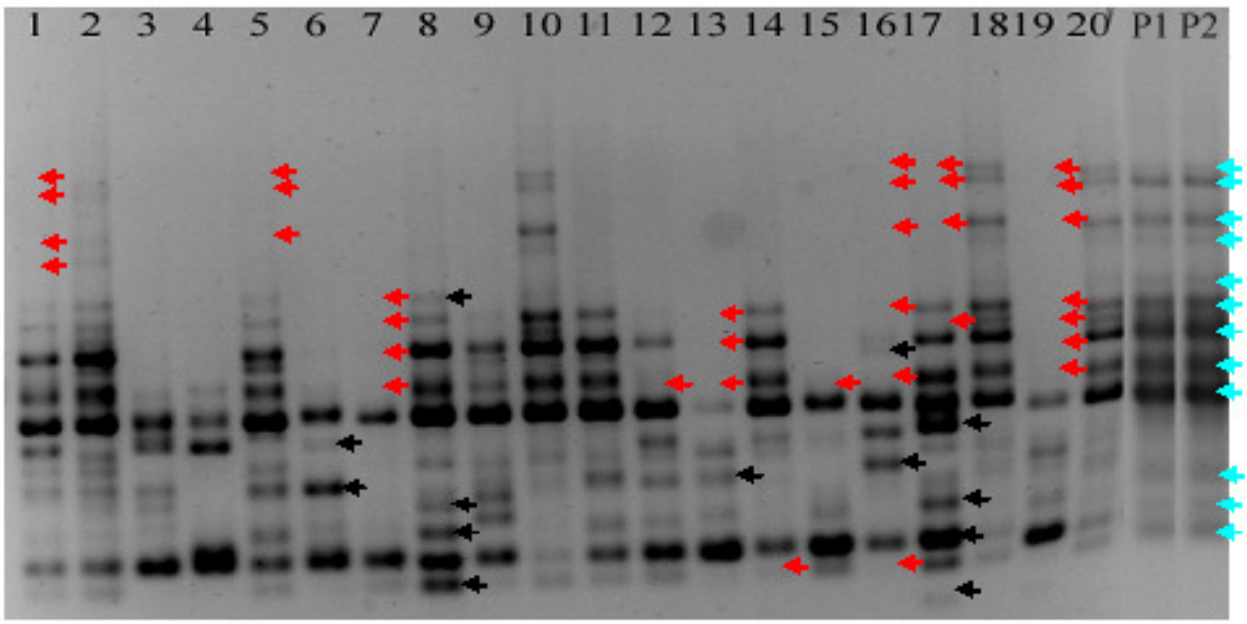


Figure 5.9: Comparative analyses of agave tissue culture accessions by IRAP with A17

IRAP analysis was carried out using retrotransposon LTR specific primer (A17 -190- IRAP). Negative of agarose gel.Lanes 1- 20 are tissue culture lines. P1 and P2 are from the parent (donor) plant. Green arrows show IRAP bands from the parent plant, red arrows show missing bands black arrows show new insertions.

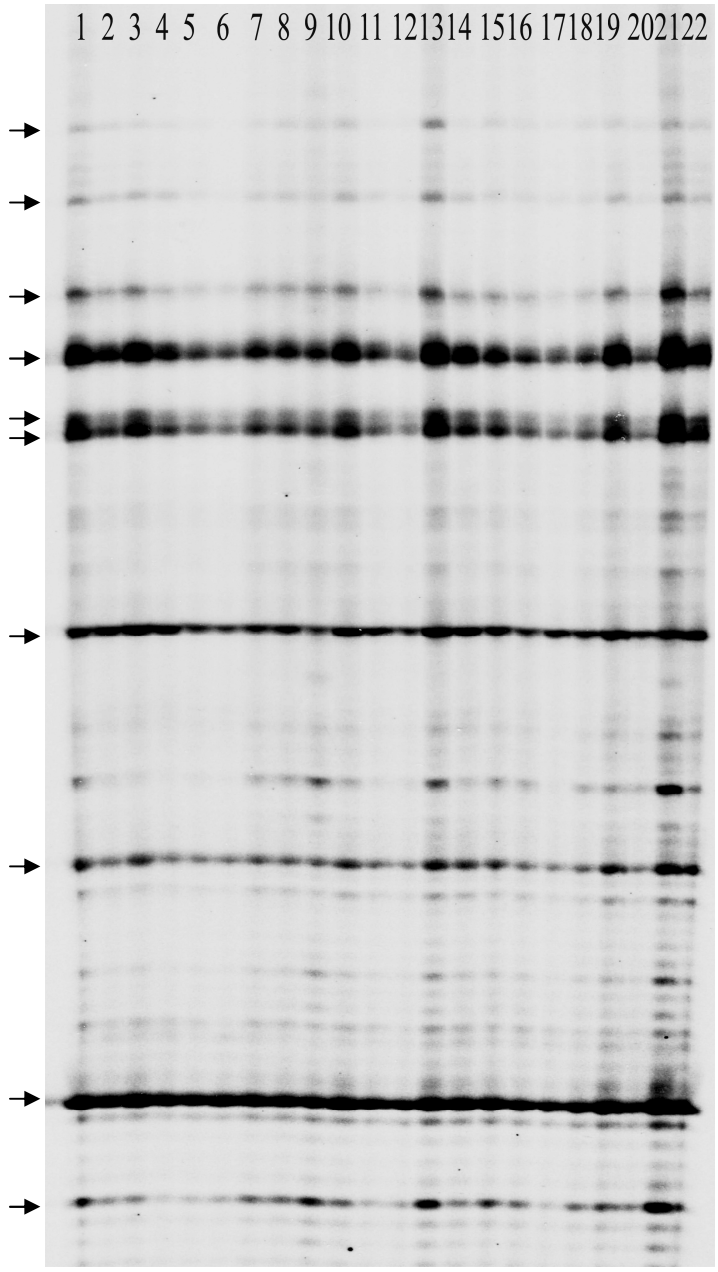


Figure 5.10**: AFLP analysis of agave tissue culture accession and parent plant.**
AFLP banding pattern of agave tissue culture accessions compared to their parent plant. Lanes 1-20 represent tissue culture accessions while lane 21 and 22 represent the parent plant. Arrows show the major bands shared by all of the tissue culture accessions as well as the parent plant. No differences can be seen from the AFLP analysis of these tissue culture accessions compared to the parent banding pattern of the parent plant.

5.2.8: Transcriptional activity of retrotransposons under the effect of tissue culture.

In order to be transpositionally active retrotransposons must produce transcripts. The previous section shows high levels of polymorphism of some classes of retroelements in the tissue culture lines which seems to be retroelement specific. This suggests that these classes of retroelements may be activated under these conditions and may therefore be producing transcripts.

If transcripts are present then this would be additional evidence that retroelement activity was the cause of some of the variations shown in the tissue culture lines and would also give some support to the hypothesis that some of the variation during vegetative propagation was caused by retroelement activity. As probes were available for a number of different classes of retroelement in agave it was possible to investigate transcriptional activity by northern blotting. The DNA sequences which have also been produced in previous work make it possible to detect retroelement-specific transcripts by RT-PCR.

Northern blotting is a molecular biology technique used for gene expression studies. The name of the technique comes from its similarity to the Southern blotting named after the scientist Edwin Southern. The major difference is that RNA is detected in northern blotting instead of DNA. Both DNA and RNA either radioactively or non-isotopically labeled probes can be used in northern blotting. The technique was developed in 1977 by James Alwin, David Kemp, and George Stark (Alwine et al., 1977) at Stanford University. In the present experiment the transcription of retrotransposons in agave tissue culture was evaluated by northern blotting using total RNA from different tissue culture lines of agave Figure 5.11, while DNA fragments belonging to reverse transcriptase of retrotransposons labeled with a non radioactive labeling agent were used as probes. However no transcriptional activity could be detected at this occasion as there were no signals on the photographic film.

As no transcripts were detected by northern blotting, it was decided to use RT- PCR (reverse transcriptase polymerase chain reaction) because RT- PCR is more sensitive than northern blotting. This is a technique in molecular biology used for the amplification of a defined piece of RNA. This technique is highly sensitive and capable of amplifying very small amount of RNA. It can also detect very small numbers of transcripts.

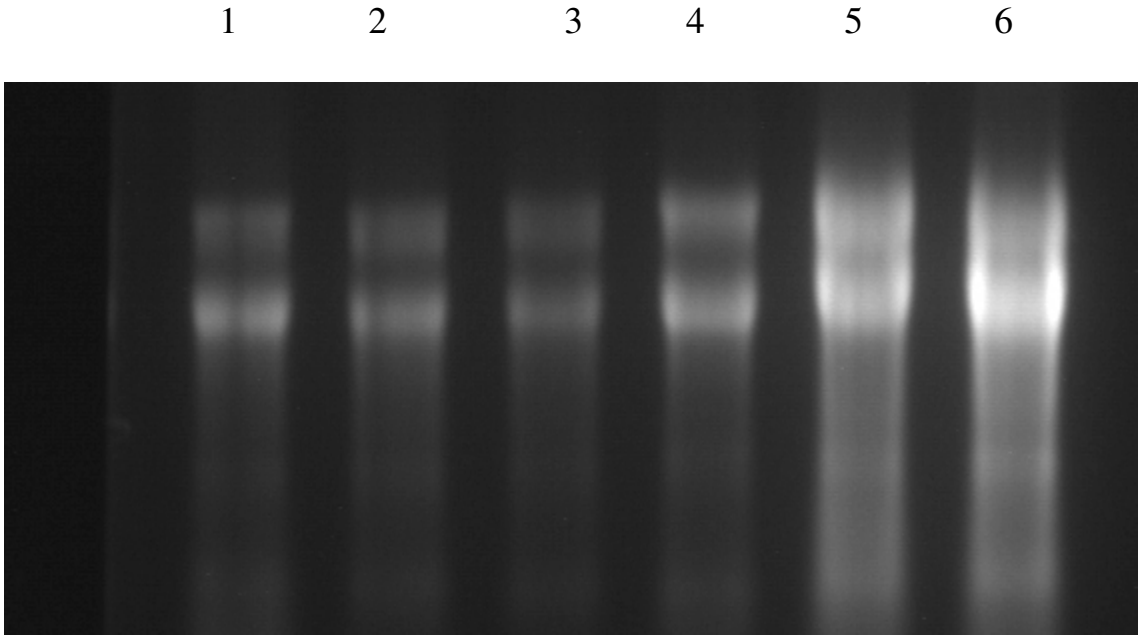


Figure: 5.11. Total RNA extracted from tissue culture linesof *Agave tequilana*
Total RNA was extracted from the tissue culture of agave and used in the northern blotting. Lanes 1-6 represent different tissue culture lines used for the extraction

An increasing use of RT- PCR to follow transcription of retrotransposons has made it possible to detect rare retrotransposons which show less dramatic increase in the level of their transcripts (Neumann et al., 2003). The single stranded RNA is first reverse transcribed into cDNA followed by the amplification of this DNA using PCR. RT- PCR has been used to evaluate the transcription of *Tto1* (Hirochika, 1993) , *Tos17* (Hirochika, 1996), *Rem1*(Ramallo et al., 2008) and retrotransposons from sweet potato (Tahara et al., 2004). In this study degenerate primers from the reverse transcriptase RT region of retrotransposons were used in RT-PCR while cDNA synthesized using total RNAs from the tissue culture of agave were used instead of DNA. The RT- PCR was positively controlled by using primers

(9F) 5'-ATGGTTTCTCTTGGCTATATGG-3'and

(13R) 5'-GCATACGTTTCATGCTTGTTTGAG -3'

These primers belong to the conserved chloroplast *ndhB* gene which is found in the inverted repeat region of chloroplast genome and is characterised across a wide range of plant taxa (Wolfetal., 1987; Goremykin et al., 1996). Primers 9F and 13R flank type II intron in the chloroplast *ndh*B (Graham and Olmstead, 2000). They amplify a 385bp fragment and have been successfully used as a positive control by Vaughan et al. (2005) in a wide variety of plants for screening purposes. In this study these primers were successfully used to amplify a 385bp fragment from agave (Figure 5.12) using cDNA synthesized from agave tissue culture. The amplification of 385bp fragment from agave confirmed the reliability of the technique and successful synthesis of cDNA from agave tissue culture. However the PCR using degenerate primers from the reverse transcriptase region of retrotransposons did not generate any products. This finding shows that no transcripts of retrotransposon could be detected in the tissue culture of agave. Unfortunately we could not detect any transcripts of retrotransposons in this study (Figure 5.12 lane 1-9).

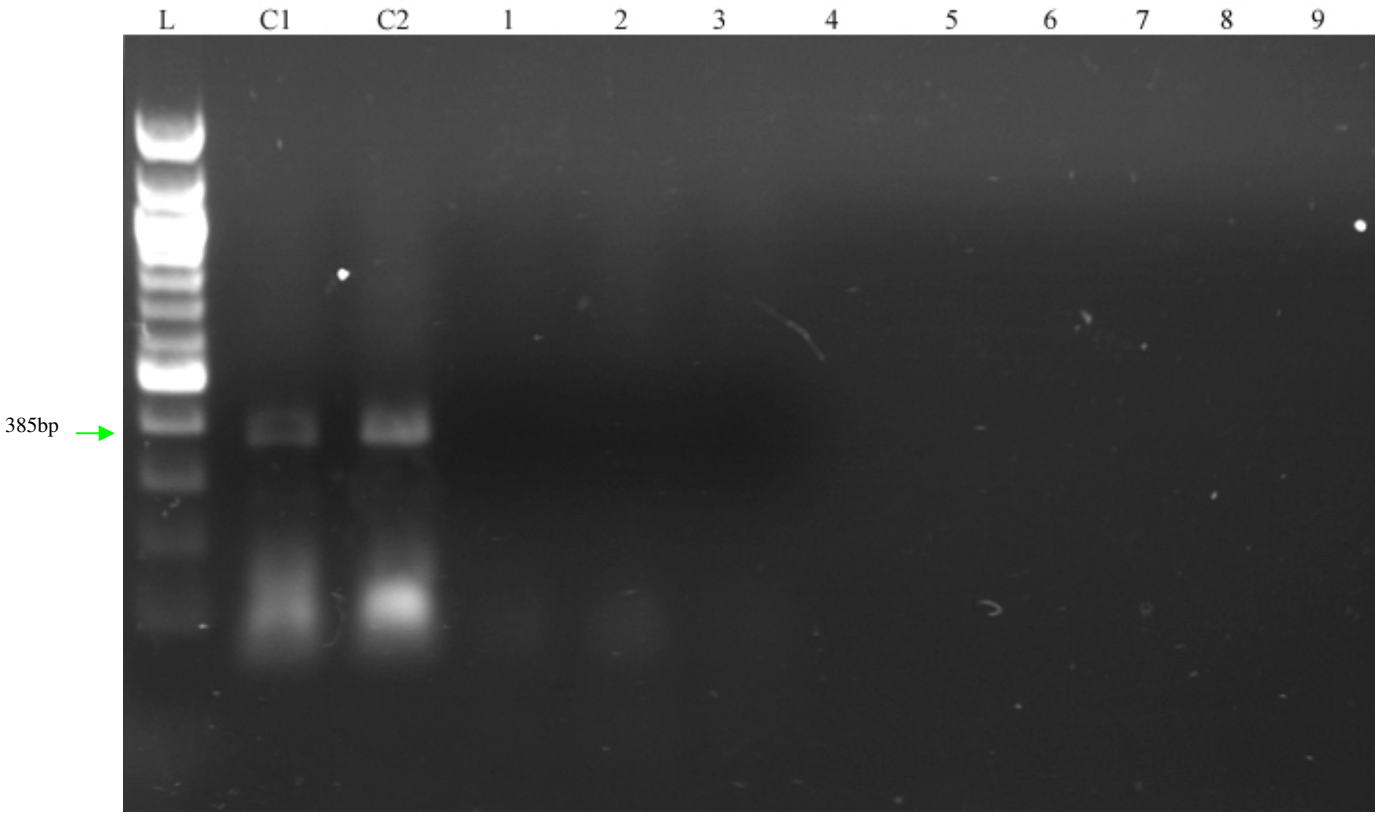


Figure 5.12: RT- PCR

RT- PCR analysis of retrotransposons transcripts in agave tissue culture was done by using cDNA synthesized form agave tissue culture. The PCR was carried out using retrotransposon RT degenerate primers for the detection of transcripts Ty1-copia retrotransposons reverse transcriptase. C1 and C2 show the positive controls using primers from conserved chloroplast *ndhB* gene. Lanes 1-9 represent RT-PCR with retrotranspososn RT degenerate primers. Green arrow represents the 385bp fragment of chloroplast *ndhB* gene.

It is clear from Figure 5.12 that there were no transcripts of *Ty1-copia* retrotransposons in the tissue culture used in this study and that the technology was working because a fragment of 385bp was amplified by the control primers. This result also confirms that there was no contamination in the cDNA used in the reaction as a bigger PCR product of 1065pb would have been amplified in case of DNA contamination with the the control primers (Vaughan et al., 2005). On the other hand DNA contamination would have resulted in the amplification of a 500bp fragment of retrotransposon reverse transcriptase with degenerate RT primers. So no transcriptional activity of retrotransposons was detected at this occasion, however this result does

not negate the presence of transcriptional activity in agave tissue culture and an evaluation of transcriptional activity at a different age or stage of culture might produce different results.

5.3: Discussion

5.3.1: Retrotransposons cause genetic instability and somaclonal variation during vegetative propagation and tissue culture of *Agave tequilana*

The aim of this chapter was to investigate asexual genetic diversity in agave as well as retrotransposon induced somaclonal variations in agave tissue culture. The experimental strategy was based on the evaluation of genetic variability among the parent and baby agave plants caused by retrotransposons during vegetative propagation. The genetic variability can be caused by insertions as well as the deletions of retrotransposons.. The genetic diversity among clones of different species of agave has been evaluated using molecular markers like AFLP and ISTR (inverse sequence tagged repeat) (Infante et al., 2006). In this work Infante and colleagues found asexual genetic diversity in *Agave tequilana* and their results contradicted the previous studies of genetic diversity in *Agave tequilana* (Vega et al., 2001).

In this study the evaluation of genetic variability among different agave plants and their baby plants was carried out. The results described here suggest that there is some genetic varability among different parent plants and their baby plants in the form of polymorphic IRAP bands. IRAP presents an opportunity to see polymorphism in the high copy number retrotransposons present in close proximity. The results described in this chapter (Figure 5.1, 5.2 and 5.3) show differences in IRAP and SSAP banding pattern. Retrotransposon polymorphism seen here is very interesting as polymorphic bands are in the form of new insertions as well as deletion of original bands.

Most of the activities of the transposable elements give rise to changes in genome structure and organization. Due to these reasons McClintock originally named them controlling elements (McClintock, 1949) and proposed that one of their major roles in evolution was to serve as a source of hypermutagenicity that could create surviving individuals from a population stressed to the point of annihilation (McClintock, 1984). So it would be impossible to have any meaningful conception of plant genome structure and evolution without understanding of the contribution of transposable elements (Bennetzen, 2000). Retrotransposons have a range of possible activities associated with alterations in the genome structure and function. Chromosome breakage and rearrangement, insertional mutation, altered gene regulation and sequence amplification are all

identified outcomes of transpositional or recombinational potential of retrotransposons (Bennetzen, 2000).

In the present study my data shows that genetic rearrangements take place during vegetative propagation of agave plants through rhizomes as described by the rearrangement of IRAP banding patterns among parent and baby agave plants. This type of mechanism is uniquely seen in agave. However agave genome might have responded to the genomic threats posed to it by continuous clonal propagation and an active retrotransposon replication might have started giving rise to new insertions. Interestingly most of these genetic rearrangements are in the form of deletions rather than new insertions. The existence of new insertion could be explained by the possibility of retrotransposon activity wich can give rise to polymorphic bands by inserting into new locations in the genome. It is proven that retrotransposons have the capacity to transpose into different parts of the genome thus causing structural and functional changes but they also appear to change their own structure much faster than genic sequences within the same genome. Deletions and internal rearrangements are common perhaps due to failed transposition events (Bennetzen, 2000) as some retrotransposons become defective during the transcription and transposition due to the occurrence of stop codons and frame shifts. Many LTR retrotransposons in plants are defective due to internal deletions, rearrangements and replacements (Hu et al., 1995). The same type of genetic rearrangement seems to be occurring in the vegetatively propagated agave baby plants resulting in the deletion of IRAP bands and genetic variability among parent and baby plants.

The movements of mobile genetic elements especially LTR retrotransposons can create a great variety of mutations in plant genomes. It was the characterization of an insertional mutant in the maize *Adh* gene that allowed the description of the first retrotransposon in plants, the Bs1 element (Johns et al., 1985). In a similar way *Tnt1*, the first active retrotransposon described in plants was also isolated after its insertion within the tobacco *Nitrate reductase* gene (Grandbastien et al., 1989). Given their mutagenic potential these elements can be a threat to the genome they reside (Kidwell and Lisch, 2000). Bennetzen and Kellogg (1997) suggested that in the absence of any vigorous mechanism to counteract this process, plant genomes would be on an inevitable road to genomic obesity. However it is now understood that plant genomes have developed mechanisms to control and reduce the mutagenic activity of retrotransposons (Hirochika et al., 2000; Vicient et al., 1999a). Among these mechanisms post transcriptional gene silencing (PTGS) and promoter inactivation by transcriptional gene silencing (TGS) are two important mechanisms that control the expression of retrotransposons in plants (Vance and Vaucheret,2002 ; Cheng et al, 2006).

Plant genomes have also evolved mechanisms like unequal homologous recombination and illegitimate recombination which can cause deletions resulting in the removal of DNA from plant genomes (Devos et al., 2002; Shirasu et al., 2000). These types of counter balancing mechanism that can create deletions in the genomes have also been reported in the insect genomes (Petrove et al, 1996). In *Arabidopsis* the most common form of deletions is represented by solo LTRs (Devos et al., 2002) , however these deletions are present in all the organisms which have LTR retrotransposons. In yeast, they have been shown to be the outcome of intrastrand homologous recombination (Roider et al, 1980). Unequal homologous recombination and illegitimate recombination are also common in rice (Ma et al., 2004), infect unequal homologous recombination is more common in rice than *Arabidopsis* as compared to other type of deletions (Ma et al., 2004). It is now well established that illegitimate recombination is the driving force behind genome size reduction by removing a large amount of genomic DNA in *Arabidopsis* than unequal homologous recombination (Devos et al, 2002). This mechanism of accumulation of small deletions is also responsible for the DNA loss in Drosophila; however it is not the predominant force responsible for the DNA loss in rice genome (Vitte et al, 2007). It can clearly be seen from the results described here in this chapter that there are some new insertions of retrotransposons exactly at the same place in unrelated plants as well as deletions from the same place in different plants. This is highly un likely that retrotransposons can insert at the same place in unrelated plants. However if a retrotransposon some how makes a new insertion in meristimatic tissue and it is fixed before this meristimatic tissue gives rise to a number of clonal plants, this might remain at the same place. *Agave tequilana* has fairly large genome (2C=8.8 pg) containing high copy number retrotransposons (chapter 4) which occur in the form of clusters of closely related heterogeneous retrotransposons (chapter 3). In the present study the presence of deletions in the form of loss of IRAP bands suggest that agave has also developed a counterbalancing mechanism to control the growth of it genome. The presence of these mechanism in *Arabidopsis* (Devos et al., 2002) and rice (Ma et al., 2004) as well as insects (Petrove et al,1996) and maize genomes (Ilic et al,2003) suggests that existence of a counterbalancing mechanism in the genome of agave like any other plant is inevitable.

Retrotransposon polymorphism has also been shown in the form of polymorphic IRAP bands in different tissue culture lines of *Agave tequilana*. Like parent and baby plants tissue culture lines are showing some new insertions along with some deletions as decribed in Figure 5.7, 5.8 and 5.9.Despite the fact that retrotransposons are abundant in the geneome of *Agave tequilana* and can be responsible for the somaclonal variations in the agave clones through their transpositional activities.one can argue that the variations among the clonally propagated agave plants and tissue

culture accessions of agave might not be related to retrotransposons at all. To test this possibility we have performed a simple AFLP experiment evaluating any general variations among the clones of agave. No variation was detected among the clones of agave by AFLP, so the genetic variability that exists in these clones has to be retrotransposon specific, high copy number and heterogeneity of retrotransposons in agave strongly supports this hypothesis, but the exact mechanism responsible for this is yet to be explained. Interestingly no transcripts of active retrotransposons were detected in the present study despite the genetic variation among clones of *Agave tequilana.* The RT-PCR as well as northern blotting did not detect any transcriptional activity at this occasion. A large scale evaluation of transcriptional activity at different stages and ages of tissue culture would probably give an idea about the transcription of retrotransposons in agave. Nevertheless agave might have developed some mechanism of its own with respect to retrotransposons to face the activity of retrotransposons.

In summary it is not clear that which particular mechanism is responsible for the genetic variability among vegetatively propagated agave plants and clones of agave produced by tissue culture. Any of the above mentioned mechanism or a combination of mechanism can be responsible for the presence of deletions as well as new insertions in the genome of agave. To answer these questions about the retrotransposon expression, stress activation and somaclonal variations, a large scale expression analysis is required. The evaluation of transcriptional activation under the effect of different stresses and comparison of parents and their clonally propagated progenies in large natural populations using different types of molecular markers would possibly be able to answer these questions. This study provides a base for the in depth understanding of retrotransposonss in agave and their role in the evolution of agave genome.

CHAPTER 6
Concluding Remarks

6.1: A brief overview of this study.

The retrotransposons isolated and characterized in this work show that the genome of *Agave tequilana* contains several families of highly abundant and heterogeneous *Ty1-copia* elements. The *Ty1-copia* retrotransposons characterized here comprises a large proportion of the agave genome but this proportion is likely to be an underestimate, as the elements isolated do not represent the *entire* population of sequences. The fact that a similar population of retrotransposon sequences that was isolated using a conserved primer from the RNaseH gene (Bousios 2008) shows that the population isolated here at best is representative of approximately 50% of the total sequences. If we also consider that this study does not estimate the numbers of the metaviridae then the complete retroelement load in *A. tequilana* may be very high indeed.

Phylogenetic analysis based on RT sequences of isolated retrotransposons revealed four major subgroups of retrotransposons which contain elements which are closely related to each other; however some individual elements with less sequence similarities to the major subgroups were also identified. Ty1-*copia* elements from *Agave tequilana* present a reverse transcriptase domain highly conserved among themselves and retrotransposon families from other plant species including rice and *Arabidopsis*. 9 elements from the isolated population were successfully used as probes for the estimation of copy number and evaluation of heterogeneity among the major subgroups. The copy number estimation and the evaluation of heterogeneity not only confirmed that the populations as a whole are high in number and heterogeneous but also that particular individual elements were also had a high copy number and highly heterogeneous. All this strongly suggests that there are far more *Ty1-copia* retrotransposons in the genome of *Agave tequilana* than were revealed in this study.

Retrotransposon based molecular marker systems such as SSAP and IRAP revealed the retroelement populations to be polymorphic among populations of clonally propagated agave plants and their vegetatively produced daughter plants. These elements were also polymorphic among tissue culture lines of *Agave tequilana.* The retrotransposon polymorphism observed here in this work gave rise to the assumption that these elements might be transcriptionally active in

agave causing insertional polymorphism among clonal plants and tissue culture lines but interestingly no transcriptional activity was detected in these tissues. There is a body of evidence presented here that the polymorphism detected is related to the retroelements as the retroelement-based marker technologies give polymorphism where methods such as AFLP which are based on restriction site polymorphism are not polymorphic. Many of the polymorphisms however relate to the loss of markers as well as the generation of new insertions so it is likely that in tissue culture the extremely high load of retroelements in Agave along with high levels of recombination may be mediating reorganization and deletion of sections of the genome. The independent generation of similar changes in independent tissue culture lines confirms this as we would expect that due to the close proximity of particular insertions some of the local sequences will be more unstable than others. It is possible that this provides a means for the generation of local variation in clonal populations of Agave which may be important for a plant which has a very long sexual generation time and may explain why these plants have tolerated, or even encouraged an extremely large retroelement load. This is an interesting area of study and it will be very interesting for future work to look in more detail at the exact nature of these changes in the clonally propagated plants.

6.2: Retrotransposons in *Agave tequilana*, prospects and possibilities

The presence of a large population of retrotransposons in *Agave tequilana* has provided an opportunity for a revised evaluation of *Agave tequilana* and its relationship with other species in the Agave genus. These elements can used as useful molecular biology tools to see both inter and intra specific relationship of plants. They can be developed into molecular markers like SSAP and IRAP which are highly informative and have been used in a wide variety of plant species including pea (Ellis et al., 1998; Pearce et al., 2000; Smykal, 2006), tomato (Tam et al., 2005), common bean (Galindo et al., 2004), maize (Garcia-Martinez and Martinez-Izquierdo, 2003), banana (Teo et al., 2005). In fact LTR retrotransposon have already been developed as molecular markers as part of another study in our lab (Bousios et al., 2007), an evaluation based on these molecular markers appeared to be challenging for the current taxonomic classification of different agave species, varieties and cultivars. It was revealed that some agaves considered as separate species were actually closely related to *Agave tequilana* and might actually be different varieties of *Agave tequilana.* This therefore has the potential of bringing other germplasm into the breeding programmes which may speed up the improvements that are needed in this crop.

The characterization of *Ty1-copia* retrotransposons here in this work identified subgroups of closely related, high copy number heterogeneous elements pointing towards bursts of retrotransposon activity in recent past. Such burst of retrotransposons activity might have played a vital role in the speciation of agaves and the shape of relatively young (8-10 Million years old) Agave genus (Good-Avila et al. 2006). The speciation in Agave genus has observed two speciation peaks, one between 8-6 Mya and the other one between 3-2.5 Mya. The first speciation peak particularly coincides with the significant rise in drought in central Mexico. It is well established that retrotransposons are activated under stresses and unusual environmental conditions (Grandbastien, 1998), so the present shape of Agave genus might actually be a result of stress activation of retrotransposons. Moreover retrotransposons have played a crucial role in the shape and evolution of eukaryotic genomes such as *Arabidopsis*, rice and barley (Devos et al., 2002; Kalendar et al., 2000; Pereira, 2004), and the large genome of *Agave tequilana* (2C = 8.8 pg) might as well an outcome of retrotransposon activity in recent past as indicated by the high copy number of *Ty1-copia* retrotranspsons estimated in this work.

As retroelements are so highly variable in agave and are also good markers for other genomic changes they are good candidates for the study of agave classification and the identification of wild relatives. Thus selection and breeding programs based on molecular data from retrotransposons of agave can help in finding resistance against insects and diseases. In this way issues around agave could be solved by saving the crop as well as improving the conditions of growers. However these breeding and crop improvement programs must be accompanied by agriculture development programs to create awareness about the importance of genetic diversity in agave in the rural communities linked with gave cultivation.

6.4: Final remarks

In the work presented here every effort has been made to achieve the goals of this research. The characterisation of *Ty1-copia* retrotransposon population of *Agave tequilana* revealed that retrotransposons are a major component of agave genome. They can be very useful in understanding the genetics of such an important commercial crop. Retrotransposons can also provide the raw material for the development of a breeding and conservation program for agave. As PhD is a long process which can give rise to many new questions and can come across new findings. The presence of unique type of retrotransposon insertions deletions can give rise to new research and improve our knowledge of clonal diversity with respect to retrotransposons. It has recently been found in our lab that the genome of *Agave tequilana* also contains a lineage of

sireviruses and in recent work I carried out a detailed analysis of clonally propagated plants with respect to sireviruses. Due to administrative deadlines and complications this work could not be included in this thesis but this work confirms that sireviruses are of extremely high copy number and are highly polymorphic in the tissue culture material. This new area of genetic research in agave would provide opportunities for genetic and evolutionary research in plants and will hopefully strengthen an industry on which so many people in Mexico depend for their livelihood and way of life.

References

Araujo, P.G., J.M. Casacuberta, A.P.P. Costa, R.Y. Hashimoto, M.A. Grandbastien, and M.A. Van Sluys. 2001. Retrolyc1 subfamilies defined by different U3 LTR regulatory regions in the Lycopersicon genus. Molecular Genetics and Genomics 266:35-41.

Arita, H.T. 1991. Spatia segregation in long - nosed bats, *Leptonycteris nivalis* and *Leptonycteris curasoae* in Maxico. Journal of Mammalogy 72:706-714.

Arizaga, S., and E. Ezcurra. 1995. Insurance against reproductive faliuer in a semelparous plant-bulbil formation in *Agave macroacantha* flowering stalks. Oecologia 101:329-334.

Arizaga, S., and E. Ezcurra. 2002. Propagation mechanisms in Agave macroacantha (Agavaceae), a tropical arid-land succulent rosette. American Journal of Botany 89:632-641.

Alves, E., I. Ballesteros, R. Linacero, and A.M. Vazquez. 2005. RYS1, a foldback transposon, is activated by tissue culture and shows preferential insertion points into the rye genome, pp. 431-436, Vol. 111. Springer.

Alwine, J.C., D.J. Kemp, and G.R. Stark. 1977. Method for Detection of Specific Rnas in Agarose Gels by Transfer to Diazobenzyloxymethyl-Paper and Hybridization with DNA Probes. Proceedings of the National Academy of Sciences of the United States of America 74:5350-5354.

Ammiraju, J.S.S., M. Luo, J.L. Goicoechea, W. Wang, D. Kudrna, C. Mueller, J. Talag, H.R. Kim, N.B. Sisneros, and B. Blackmon. 2006. The Oryza bacterial artificial chromosome library resource: construction and analysis of 12 deep-coverage large-insert BAC libraries that represent the 10 genome types of the genus Oryza, pp. 140, Vol. 16. Cold Spring Harbor Lab.

Avramova, Z., P. SanMiguel, E. Georgieva, and J.L. Bennetzen. 1995. Matrix attachment regions and transcribed sequences within a long chromosomal continuum containing maize Adh1, pp. 1667, Vol. 7. Am Soc Plant Biol.

Bajaj, Y.P.S. 1990. Cryopreservation of Germplasm of Vegetatively Propagated Crops. Bulletin De La Societe Botanique De France-Actualites Botaniques 137:99-114.

Banga, S.S., A. Velazquez, and J.B. Boyd. 1991. P transposition in Drosophila provides a new tool for analyzing postreplication repair and double-strand break repair, pp. 79-88, Vol. 255. Elsevier.

Barow, M., and A. Meister. 2002. Lack of correlation between AT frequency and genome size in higher plants and the effect of nonrandomness of base sequences on dye binding. Cytometry 47:1-7.

Bernstein, E., A.A. Caudy, S.M. Hammond, and G.J. Hannon. 2001. Role for a bidentate ribonuclease in the initiation step of RNA interference, pp. 363-366, Vol. 409. Nature Publishing Group.

Bousios, A., I. Saldana-Oyarzabal, A.G. Valenzuela-Zapata, C. Wood, and S.R. Pearce. 2007. Isolation and characterization of Ty1-copia retrotransposon sequences in the blue agave (*Agave tequilana* Weber var. azul) and their development as SSAP markers for phylogenetic analysis. Plant Science 172:291-298.

Brandes, A., H. Thompson, C. Dean, and J.S. HeslopHarrison. 1997. Multiple repetitive DNA sequences in the paracentromeric regions of *Arabidopsis* thaliana L. Chromosome Research 5:238-246.

Britten, R.J., and D.E. Kohne. 1968. Repeated Sequences in DNA. Science 161:529-&.

Brookfield, J.F.Y. 2005. The ecology of the genome - Mobile DNA elements and their hosts. Nature Reviews Genetics 6:128-136.

Bureau, T.E., and S.R. Wessler. 1994. Stowaway: a new family of inverted repeat elements associated with the genes of both monocotyledonous and dicotyledonous plants, pp. 907, Vol. 6. Am Soc Plant Biol.
Badano, E.I., and F.I. Pugnaire. 2004. Invasion of Agave species (Agavaceae) in southeast Spain: invader demographic parameters and impacts on native species. Diversity And Distributions 10: 493-500.
Batzer, M.A., and P.L. Deininger. 2002. Alu repeats and human genomic diversity. Nature Reviews Genetics 3:370-379.
Beguiristain, T., M.A. Grandbastien, P. Puigdomenech, and J.M. Casacuberta. 2001. Three Tnt1 subfamilies show different stress-associated patterns of expression in tobacco. Consequences for retrotransposon control and evolution in plants. Plant Physiology 127:212-221.
Bennett, M.D., and I.J. Leitch. 1995. Nuclear DNA amounts in angiosperms. Annals of Botany 76:113-176.
Bennett, M.D., and I.J. Leitch. 2004. Angiosperm DNA C-values database (release 5.0, Dec. 2004). http://www.kew.org/cvalues/homepage.html.
Bennetzen, J.L. 1996. The contributions of retroelements to plant genome organization, function and evolution. Trends in Microbiology 4:347-353.
Bennett, M.D., and I.J. Leitch. 2005. Plant genome size research: A field in focus. Annals of Botany 95:1-6.
Bennetzen, J.L. 2000. Transposable element contributions to plant gene and genome evolution. Plant Molecular Biology 42:251-269.
Bennetzen, J.L. 2002. Mechanisms and rates of genome expansion and contraction in flowering plants. Genetica 115:29-36.
Bennetzen, J.L. 2005. Transposable elements, gene creation and genome rearrangement in flowering plants. Current Opinion in Genetics & Development 15:621-627.
Bennetzen, J.L., and E.A. Kellogg. 1997. Do plants have a one-way ticket to genomic obesity? Plant Cell 9:1509-1514.
Bennetzen, J.L., J.X. Ma, and K. Devos. 2005. Mechanisms of recent genome size variation in flowering plants. Annals of Botany 95:127-132.
Bentham, G., and J.D. Hooker. 1883. Genera plantarum. Vol. III, part 2. London: Reeve.
Boeke, J.D., and V.G. Corces. 1989. Transcription and Reverse Transcription of Retrotransposons. Annual Review of Microbiology 43:403-434.
Boeke, J.D., T.H. Eickbush, S.B. Sandmeyer, and D.F. Voytas. 2006. Index of Viruses - Pseudoviridae (2006). In: ICTVdB - The Universal Virus Database, version 4. Buchen-Osmond, C (Ed). Columbia University, New York, USA http://www.ncbi.nlm.nih.gov/ICTVdb/Ictv/fs_index.htm.
Bogler, D.J., and B.B. Simpson. 1995. A Chloroplast DNA study of the *Agavaceae*. Systematic Botany 20:191-205.
Boom, A., J.S.S. Damste, and J.W. de Leeuw. 2005. Cutan, a common aliphatic biopolymer in cuticles of drought-adapted plants. Organic Geochemistry 36:595-601.
Casacuberta, J.M., and M.A. Grandbastien. 1993. Characterization of Ltr Sequences Involved in the Protoplast Specific Expression of the Tobacco Tnt1 Retrotransposon. Nucleic Acids Research 21:2087-2093.
Casacuberta, J.M., and N. Santiago. 2003. Plant LTR-retrotransposons and MITEs: control of transposition and impact on the evolution of plant genes and genomes. Gene 311:1-11.
Casacuberta, J.M., S. Vernhettes, and M.A. Grandbastien. 1995. Sequence Variability within the Tobacco Retrotransposon Tnt1 Population. Embo Journal 14:2670-2678.
Casacuberta, J.M., S. Vernhettes, C. Audeon, and M.A. Grandbastien. 1997. Quasispecies in retrotransposons: a role for sequence variability in Tnt1 evolution. Genetica 100:109-117.

Cavalier-Smith, T. 1978. Nuclear volume control by nucleoskeletal DNA, selection for cell volume and cell growth rate, and solution of DNA C - value paradox. Journal of Cell Science 34:247-278.
Chavanne, F., D.X. Zhang, M.F. Liaud, and R. Cerff. 1998. Structure and evolution of Cyclops: a novel giant retrotransposon of the Ty3/Gypsy family highly amplified in pea and other legume species. Plant Molecular Biology 37:363-375.
Chenna, R., H. Sugawara, T. Koike, R. Lopez, T.J. Gibson, D.G. Higgins, and J.D. Thompson. 2003. Multiple sequence alignment with the Clustal series of programs. Nucleic Acids Research 31:3497-3500.
Costa, A.P.P., K.C. Scortecci, R.Y. Hashimoto, P.G. Araujo, M.A. Grandbastien, and M.A. Van Sluys. 1999. Retrolyc1-1, a member of the Tnt1 retrotransposon super-family in the Lycopersicon peruvianum genome. Genetica 107:65-72.
Colunga, G.P., and F. Maypat. 1993. AGAVE STUDIES IN YUCATAN, MEXICO .1. PAST AND PRESENT GERMPLASM DIVERSITY AND USES. Economic Botany 47:312-327.
Chalker, D.L., and S.B. Sandmeyer. 1992. Ty3 Integrates within the Region of Rna Polymerase-Iii Transcription Initiation. Genes & Development 6:117-128.
Cronquist, A. 1981. An integrated system of classification of flowering plants. New York: Columbia Univ. Press.
Dahlgren, R.M., and P.F. Yeo. 1985. The families of the monocotyledons. Berlin: Springer-Verlag.
Capy, P., G. Gasperi, C. Biemont, and C. Bazin. 2000. Stress and transposable elements: co-evolution or useful parasites? Heredity 85:101-106.
Chaw, S.M., C.C. Chang, H.L. Chen, and W.H. Li. 2004. Dating the monocot–dicot divergence and the origin of core eudicots using whole chloroplast genomes, pp. 424-441, Vol. 58. Springer.
Colunga, G.P., and F. Maypat. 1993. Agave studies in yucatan, mexico .1. Past and present germplasm diversity and uses. Economic Botany 47:312-327.
Craig, N.L. 2002. Mobile DNA II Amer Society for Microbiology.
Deininger, P.L., J.V. Moran, M.A. Batzer, and H.H. Kazazian. 2003. Mobile elements and mammalian genome evolution. Current Opinion in Genetics & Development 13:651-658.
Deng, H., V.A. Bloomfield, J.M. Benevides, and G.J. Thomas. 1999. Dependence of the raman signature of genomic B-DNA on nucleotide base sequence. Biopolymers 50:656-666.
Deragon, J.M., N. Gilbert, L. Rouquet, A. Lenoir, P. Arnaud, and G. Picard. 1996. A transcriptional analysis of the S1(Bn), (Brassica napus) family of SINE retroposons. Plant Molecular Biology 32:869-878.
Dalton, R. 2005. Saving the agave. Nature 438:1070-1071.
Demey, J.R., E. Gamez, S. Molina, and D. Infante. 2004. Comparative study of the discriminating capacity of AFLP and ISTR markers for genetic analysis of Agave fourcroydes. Plant Molecular Biology Reporter 22:29-35.
Devos, K.M., J.K.M. Brown, and J.L. Bennetzen. 2002. Genome size reduction through illegitimate recombination counteracts genome expansion in *Arabidopsis*. Genome Research 12:1075-1079.
Ding, Y., Y.Y. Chen, D.Z. Wang, and C.R. Yang. 1989. Steroidal saponins from a cultivated form of *Agave sisalana*. Phytochemistry 28:2787-2791.
DISCUS. 2008. U.S. Tequila Market at a glance. Distilled Spirits Council of the United States.
Dodd, A.N., A.M. Borland, R.P. Haslam, H. Griffiths, and K. Maxwell. 2002. Crassulacean acid metabolism: plastic, fantastic. Journal of Experimental Botany 53:569-580.
Duvall, M.R., M.T. Clegg, M.W. Chase, W.D. Clark, W.J. Kress, H.G. Hills, L.E. Eguiarte, J.F. Smith, B.S. Gaut, E.A. Zimmer, and G.H. Learn. 1993. Phylogenetic hypotheses for the

monocotyledons constructed from RBCL sequencing data. Annals of the Missouri Botanical Garden 80:607-619.
Evans, P.D. 1984. The role of cyclic nucleotides and calcium in the mediation of the modulatory effects of octopamine on locust skeletal muscle, pp. 325, Vol. 348. Physiological Soc.
Eguiarte, L., V. Souza, and A. Silva-Montellano. 2000. Boletin de la Sociedad. Botanica de Mexico 66, 131-150.
Eickbush, T.H., and V.K. Jamburuthugoda. 2008. The diversity of retrotransposons and the properties of their reverse transcriptases. Virus Research 134:221-234.
Ellis, T.H.N., S.J. Poyser, M.R. Knox, A.V. Vershinin, and M.J. Ambrose. 1998. Polymorphism of insertion sites of Ty1-copia class retrotransposons and its use for linkage and diversity analysis in pea. Molecular and General Genetics 260:9-19.
Engler, A., and K. Prantl. 1888. Dia naturlichen Pflanzenfamilien. Teil 2, Abteil 5. Liepzig.
Feschotte, C., and S.R. Wessler. 2001. Treasures in the attic: Rolling circle transposons discovered in eukaryotic genomes. Proceedings of the National Academy of Sciences of the United States of America 98:8923-8924.
Feschotte, C., N. Jiang, and S.R. Wessler. 2002. Plant transposable elements: Where genetics meets genomics. Nature Reviews Genetics 3:329-341.
Feschotte, C., and E.J. Pritham. 2005. Non-mammalian c-integrases are encoded by giant transposable elements. Trends in Genetics 21:551-552.
Feschotte, C., and E.J. Pritham. 2007. DNA transposons and the evolution of eukaryotic genomes. Annual Review of Genetics 41:331-368.
Feng, Q., Y. Zhang, P. Hao, S. Wang, G. Fu, Y. Huang, Y. Li, J. Zhu, Y. Liu, and X. Hu. 2002. Sequence and analysis of rice chromosome 4, pp. 316-320, Vol. 420. Nature Publishing Group.
Flavell, A.J., D.B. Smith, and A. Kumar. 1992a. Extreme Heterogeneity of Ty1-Copia Group Retrotransposons in Plants. Molecular & General Genetics 231:233-242.
Flavell, A.J., E. Dunbar, R. Anderson, S.R. Pearce, R. Hartley, and A. Kumar. 1992b. Ty1-Copia Group Retrotransposons Are Ubiquitous and Heterogeneous in Higher-Plants. Nucleic Acids Research 20:3639-3644.
Flavell, R.B., M.D. Bennett, J.B. Smith, and D.B. Smith. 1974. Genome Size and Proportion of Repeated Nucleotide-Sequence DNA in Plants. Biochemical Genetics 12:257-269.
Francki, M.G. 2001. Identification of Bilby, a diverged centromeric Ty1-copia retrotransposon family from cereal rye (Secale cereale L.). Genome 44:266-274.
Friesen, N., A. Brandes, and J.S. Heslop-Harrison. 2001. Diversity, origin, and distribution of retrotransposons (gypsy and copia) in conifers. Molecular Biology and Evolution 18:1176-1188.
Gabriel, A., M. Willems, E.H. Mules, and J.D. Boeke. 1996. Replication infidelity during a single cycle of Ty1 retrotransposition. Proceedings of the National Academy of Sciences of the United States of America 93:7767-7771.
Galindo, L.M., E. Gaitan-Solis, P. Baccam, and J. Tohme. 2004. Isolation and characterization of RNase LTR sequences of Ty1-copia retrotransposons in common bean (Phaseolus vulgaris L.). Genome 47:84-95.
Garber, K., I. Bilic, O. Pusch, J. Tohme, A. Bachmair, D. Schweizer, and V. Jantsch. 1999. The Tpv2 family of retrotransposons of Phaseolus vulgaris: structure, integration characteristics, and use for genotype classification. Plant Molecular Biology 39:797-807.
Garcia-Martinez, J., and J.A. Martinez-Izquierdo. 2003. Study on the evolution of the Grande retrotransposon in the Zea genus. Molecular Biology and Evolution 20:831-841.
Garcia-Mendoza, A., and F. Chiang. 2003. The confusion of Agave vivipara L. and A. angustifolia Haw., two distinct taxa. Brittonia 55:82-87.

Gentry, H.S. 1982. Agaves from Continental North America. The University of Arizona Press, Tucson:670pp.
Gonzalez, G., S. Aleman, and D. Infante. 2003. Asexual genetic variability in Agave fourcroydes II: selection among individuals in a clonally propagated population. Plant Science 165:595-601.
Good-Avila, S.V., V. Souza, B.S. Gaut, and L.E. Eguiarte. 2006. Timing and rate of speciation in Agave (Agavaceae). Proceedings of the National Academy of Sciences of the United States of America 103:9124-9129.
Grandbastien, M., C. Audeon, E. Bonnivard, J.M. Casacuberta, B. Chalhoub, A.P.P. Costa, Q.H. Le, D. Melayah, M. Petit, C. Poncet, S.M. Tam, M.A. Van Sluys, and C. Mhiri. 2005. Stress activation and genomic impact of Tnt1 retrotransposons in Solanaceae. Cytogenetic and Genome Research 110:229-241.
Grandbastien, M.A., A. Spielmann, and M. Caboche. 1989. Tnt1, a Mobile Retroviral-Like Transposable Element of Tobacco Isolated by Plant-Cell Genetics. Nature 337:376-380.
Grandbastien, M.A. 1992. Retroelements in Higher-Plants. Trends in Genetics 8:103-108.
Grandbastien, M.A. 1998. Activation of plant retrotransposons under stress conditions. Trends in Plant Science 3:181-187.
Grandbastien, M.A., H. Lucas, J.B. Morel, C. Mhiri, S. Vernhettes, and J.M. Casacuberta. 1997. The expression of the tobacco Tnt1 retrotransposon is linked to plant defense responses. Genetica 100:241-252.
Gribbon, B.M., S.R. Pearce, R. Kalendar, A.H. Schulman, L. Paulin, P. Jack, A. Kumar, and A.J. Flavell. 1999. Phylogeny and transpositional activity of Ty1-copia group retrotransposons in cereal genomes. Molecular and General Genetics 261:883-891.
Hamilton, A., O. Voinnet, L. Chappell, and D. Baulcombe. 2002. Two classes of short interfering RNA in RNA silencing, pp. 4671-4679, Vol. 21. Nature Publishing Group.
Havecker, E.R., X. Gao, and D.F. Voytas. 2004. The diversity of LTR retrotransposons.Genome Biology 5.
Hernandez, P., R. de la Rosa, L. Rallo, A. Martin, and G. Dorado. 2001. First evidence of a retrotransposon-like element in olive (Olea europaea): implications in plant variety identification by SCAR-marker development. Theoretical and Applied Genetics 102:1082-1087.
Hendry, G.A.F. 1993.Evolutionary origins and natural functions of fructans- a climatological, biogeographic and mehanistic appraisal. New Phytologist 123:3-14.
Heslop-Harrison, J.S., A. Brandes, S. Taketa, T. Schmidt, A.V. Vershinin, E.G. Alkhimova, A. Kamm, R.L. Doudrick, T. Schwarzacher, A. Katsiotis, S. Kubis, A. Kumar, S.R. Pearce, A.J. Flavell, and G.E. Harrison. 1997. The chromosomal distributions of Ty1-copia group retrotransposable elements in higher plants and their implications for genome evolution. Genetica 100:197-204.
Hidalgo, I.J., G.V. Calleros, O.M. Vega, G. Vandemark, and V.O. Portugal. 2004. Identification and characterisation of bacteria causing soft-rot in *Agave tequilana*. European Journal of Plant Pathology 110: 317–331, 2004.
Hirochika, H. 1997. Retrotransposons of rice: their regulation and use for genome analysis. Plant Molecular Biology 35:231-240.
Hirochika, H. 2001. Contribution of the Tos17 retrotransposon to rice functional genomics. Current Opinion in Plant Biology 4:118-122.
Hirochika, H., A. Fukuchi, and F. Kikuchi. 1992. Retrotransposon Families in Rice. Molecular & General Genetics 233:209-216.
Hirochika, H. 1993. Activation of Tobacco Retrotransposons During Tissue-Culture. Embo Journal 12:2521-2528.

Hirochika, H., H. Okamoto, and T. Kakutani. 2000. Silencing of retrotransposons in *Arabidopsis* and reactivation by the ddm1 mutation. Plant Cell 12:357-368.
Hirochika, H., K. Sugimoto, Y. Otsuki, H. Tsugawa, and M. Kanda. 1996a. Retrotransposons of rice involved in mutations induced by tissue culture. Proceedings of the National Academy of Sciences of the United States of America 93:7783-7788.
Hirochika, H., H. Otsuki, M. Yoshikawa, Y. Otsuki, K. Sugimoto, and S. Takeda. 1996b. Autonomous transposition of the tobacco retrotransposon Tto1 in rice. Plant Cell 8:725-734.
Holligan, D., X.Y. Zhang, N. Jiang, E.J. Pritham, and S.R. Wessler. 2006. The transposable element landscape of the model legume Lotus japonicus. Genetics 174:2215-2228.
Hutchinson, J. 1934. The families of flowering plants. Vol. II, Monocotyledons. Oxford: Clarendon Press.
Hu, W.M., O.P. Das, and J. Messing. 1995. Zeon-1, a Member of a New Maize Retrotransposon Family. Molecular & General Genetics 248:471-480.
Infante, D., G. Gonzalez, L. Peraza-Echeverria, and M. Keb-Llanes. 2003. Asexual genetic variability in Agave fourcroydes. Plant Science 164:223-230.
Infante, D., S. Molina, J.R. Demey, and E. Gamez. 2006. Asexual genetic variability in Agavaceae determined with inverse sequence-tagged repeats and amplification fragment length polymorphism analysis. Plant Molecular Biology Reporter 24:205-217.
Infante, D., M. Osorio, S. Molina, and G. Gonzalez. 2007. Genetic improvement of asexually propagated plants, p. 721-727, *In* R. E. Litz and R. Scorza, eds. Proceedings of the International Symposium on Biotechnology of Temperate Fruit Crops and Tropical Species.
Jaligot, E., A. Rival, T. Beule, S. Dussert, and J.L. Verdeil. 2000. Somaclonal variation in oil palm (Elaeis guineensis Jacq.): the DNA methylation hypothesis, pp. 684-690, Vol. 19. Springer.
Jensen, S., M.P. Gassama, and T. Heidmann. 1999. Taming of transposable elements by homology-dependent gene silencing, pp. 209-212, Vol. 21. Nature Publishing Group.
Jovtchev, G., M. Barow, A. Meister, and I. Schubert. 2007. Impact of environmental and endogenous factors on endopolyploidization in angiosperms. Environmental and Experimental Botany 60:404-411.
Johns, M.A., J. Mottinger, and M. Freeling. 1985. A Low Copy Number, Copia-Like Transposon in Maize. Embo Journal 4:1093-1101.
Jiang, N., Z.R. Bao, X.Y. Zhang, S.R. Eddy, and S.R. Wessler. 2004. Pack-MULE transposable elements mediate gene evolution in plants. Nature 431:569-573.
Kalendar, R., T. Grob, M. Regina, A. Suoniemi, and A. Schulman. 1999. IRAP and REMAP: two new retrotransposon-based DNA fingerprinting techniques. Theoretical and Applied Genetics 98:704-711.
Kalendar, R., J. Tanskanen, S. Immonen, E. Nevo, and A.H. Schulman. 2000. Genome evolution of wild barley (Hordeum spontaneum) by BARE-1 retrotransposon dynamics in response to sharp microclimatic divergence. Proceedings of the National Academy of Sciences of the United States of America 97:6603-6607.
Kalendar, R., C.M. Vicient, O. Peleg, K. Anamthawat-Jonsson, A. Bolshoy, and A.H. Schulman. 2004. Large retrotransposon derivatives: Abundant, conserved but nonautonomous retroelements of barley and related genomes. Genetics 166:1437-1450.
Kapitonov, V.V., and J. Jurka. 2001. Rolling-circle transposons in eukaryotes. Proceedings of the National Academy of Sciences of the United States of America 98:8714-8719.
Kapitonov, V.V., and J. Jurka. 2006. New kids on the block: Self-synthesizing DNA transposons. Proceedings of the Fifth International Conference on Bioinformatics of Genome Regulation and Structure, Vol 3:167-170.

Kaeppler, S.M., H.F. Kaeppler, and Y. Rhee. 2000. Epigenetic aspects of somaclonal variation in plants, pp. 179-188, Vol. 43. Springer.
Karp, A. 1995. Somaclonal variation as a tool for crop improvement, pp. 295-302, Vol. 85. Springer.
Keulen, W., N.K. Back, A. Van Wijk, C.A. Boucher, and B. Berkhout. 1997. Initial appearance of the 184Ile variant in lamivudine-treated patients is caused by the mutational bias of human immunodeficiency virus type 1 reverse transcriptase, pp. 3346, Vol. 71. Am Soc Microbiol.
Knoop, V., M. Unseld, J. Marienfeld, P. Brandt, S. Sunkel, H. Ullrich, and A. Brennicke. 1996. copia-, gypsy and LINE-like retrotransposon fragments in the mitochondrial genome of *Arabidopsis* thaliana. Genetics 142:579-585.
Konieczny, A., D.F. Voytas, M.P. Cummings, and F.M. Ausubel. 1991. A superfamily of *Arabidopsis* thaliana retrotransposons, pp. 801, Vol. 127. Genetics Soc America.

Katz, R.A., and A.M. Skalka. 1994. The Retroviral Enzymes. Annual Review of Biochemistry 63:133-173.
Kazazian, H.H. 2004. Mobile elements: Drivers of genome evolution. Science 303:1626-1632.
Kidwell, M.G. 2002. Transposable elements and the evolution of genome size in eukaryotes. Genetica 115:49-63.
Kidwell, M.G., and D. Lisch. 1997. Transposable elements as sources of variation in animals and plants. Proceedings of the National Academy of Sciences of the United States of America 94:7704-7711.
Kidwell, M.G., and D.R. Lisch. 2000. Transposable elements and host genome evolution. Trends in Ecology & Evolution 15:95-99.
Konieczny, A., D.F. Voytas, M.P. Cummings, and F.M. Ausubel. 1991. A Superfamily of *Arabidopsis*-Thaliana Retrotransposons. Genetics 127:801-809.
Knight, C.A., N.A. Molinari, and D.A. Petrov. 2005. The large genome constraint hypothesis: Evolution, ecology and phenotype. Annals of Botany 95:177-190.
Kumar, A. 1996. The adventures of the Ty1-copia group of retrotransposons in plants. Trends in Genetics 12:41-43.
Kumar, A., S.R. Pearce, K. McLean, G. Harrison, J.S. Heslop-Harrison, R. Waugh, and A.J. Flavell. 1997. The Ty1-copia group of retrotransposons in plants: genomic organisation, evolution, and use as molecular markers. Genetica 100:205-217.
Kumar, A., and J.L. Bennetzen. 1999. Plant retrotransposons. Annual Review of Genetics 33:479-532.
Kunze, R., H. Saedler, and W.E. Lonnig. 1997. Plant transposable elements, p. 331-470 Advances in Botanical Research, Vol 27, Vol. 27. Academic Press Ltd, London.
Leeton, P.R.J., and D.R. Smyth. 1993. An Abundant Line-Like Element Amplified in the Genome of Lilium-Speciosum. Molecular & General Genetics 237:97-104.
Linares, C., A. Serna, and A. Fominaya. 1999. Chromosomal organization of a sequence related to LTR-like elements of Ty1-copia retrotransposons in Avena species. Genome 42:706-713.
Linares, C., Y. Loarce, A. Serna, and A. Fominaya. 2001. Isolation and characterization of two novel retrotransposons of the Ty1-copia group in oat genomes. Chromosoma 110:115-123.
Larkin, P.J., and W.R. Scowcroft. 1981. Somaclonal variation—a novel source of variability from cell cultures for plant improvement, pp. 197-214, Vol. 60. Springer.
Leprince, A.S., M.A. Grandbastien, and C. Meyer. 2001. Retrotransposons of the Tnt1B family are mobile in Nicotiana plumbaginifolia and can induce alternative splicing of the host gene upon insertion, pp. 533-541, Vol. 47. Springer.

Li, X., X. Yu, N. Wang, Q. Feng, Z. Dong, L. Liu, J. Shen, and B. Liu. 2007. Genetic and epigenetic instabilities induced by tissue culture in wild barley (Hordeum brevisubulatum (Trin.) Link), pp. 153-168, Vol. 90. Springer.
Liu, K., and S. Somerville. 1996. Cloning and characterization of a highly repeated DNA sequence in Hordeum vulgare L, pp. 1159, Vol. 39.
Llave, C., K.D. Kasschau, M.A. Rector, and J.C. Carrington. 2002. Endogenous and silencing-associated small RNAs in plants, pp. 1605, Vol. 14. Am Soc Plant Biol.
Lucas, H., F. Feuerbach, K. Kunert, M.A. Grandbastien, and M. Caboche. 1995. Rna-Mediated Transposition of the Tobacco Retrotransposon Tnt1 in *Arabidopsis*-Thaliana. Embo Journal 14:2364-2373.
Marillonnet, S., and S.R. Wessler. 1997. Retrotransposon insertion into the maize waxy gene results in tissue-specific RNA processing, pp. 967, Vol. 9. Am Soc Plant Biol.
Martienssen, R.A., and V. Colot. 2001. DNA methylation and epigenetic inheritance in plants and filamentous fungi, pp. 1070, Vol. 293. AAAS.
Matzke, M., A.J.M. Matzke, and J.M. Kooter. 2001. RNA: Guiding gene silencing. Science 293:1080-1083.
Mochizuki, K., M. Umeda, H. Ohtsubo, and E. Ohtsubo. 1992. Characterization of a plant SINE, p-SINE1, in rice genomes, pp. 155-166, Vol. 67. J-STAGE.
Ma, J.X., K.M. Devos, and J.L. Bennetzen. 2004. Analyses of LTR-retrotransposon structures reveal recent and rapid genomic DNA loss in rice. Genome Research 14:860-869.
Malik, H.S. 2005. Ribonuclease H evolution in retrotransposable elements. Cytogenetic and Genome Research 110:392-401.
Malik, H.S., and T.H. Eickbush. 1999. Modular evolution of the integrase domain in the Ty3/Gypsy class of LTR retrotransposons. Journal of Virology 73:5186-5190.
Malik, H.S., and T.H. Eickbush. 2001. Phylogenetic analysis of ribonuclease H domains suggests a late, chimeric origin of LTR retrotransposable elements and retroviruses. Genome Research 11:1187-1197.
Malik, H.S., S. Henikoff, and T.H. Eickbush. 2000. Poised for contagion: Evolutionary origins of the infectious abilities of invertebrate retroviruses. Genome Research 10:1307-1318.
Manninen, I., and A.H. Schulman. 1993. Bare-1, a Copia-Like Retroelement in Barley (Hordeum-Vulgare L). Plant Molecular Biology 22:829-846.
Matsuoka, Y., and K. Tsunewaki. 1999. Evolutionary dynamics of Ty1-copia group retrotransposons in grass shown by reverse transcriptase domain analysis. Molecular Biology and Evolution 16:208-217.
McCarthy, E.M., J. Liu, G. Lizhi, and J.F. McDonald. 2002. Long terminal repeat retrotransposons of Oryza sativa. Genome Biology 3.
McClintock, B. 1951. Chromosome organization and genetic expression. Cold Spring Harbor Symposia on Quantitative Biology 16:13-47.
McClintock, B. 1984. The Significance of Responses of the Genome to Challenge. Science 226:792-801.
McKelvey, S.D., and K. Sax. 1933. Taxonomic and cytological relationships of Yucca and Agave. Journal of the Arnold Arboretum 14: 76-81.
Mhiri, C., P. De Wit, and M.A. Grandbastien. 1999. Activation of the promoter of the Tnt1 retrotransposon in tomato after inoculation with the fungal pathogen Cladosporium fulvum. Molecular Plant-Microbe Interactions 12:592-603.
Mhiri, C., J.B. Morel, S. Vernhettes, J.M. Casacuberta, H. Lucas, and M.A. Grandbastien. 1997. The promoter of the tobacco Tnt1 retrotransposon is induced by wounding and by abiotic stress. Plant Molecular Biology 33:257-266.

Melayah, D., E. Bonnivard, B. Chalhoub, C. Audeon, and M.A. Grandbastien. 2001. The mobility of the tobacco Tnt1 retrotransposon correlates with its transcriptional activation by fungal factors. Plant Journal 28:159-168.
Meyers, B.C., S.V. Tingley, and M. Morgante. 2001. Abundance, distribution, and transcriptional activity of repetitive elements in the maize genome. Genome Research 11:1660-1676.
Miller, M., and K. Taube. 1993. The Gods and Symbols of Ancient Mexico and the Maya: An Illustrated Dictionary of Mesoamerican Religion. London: Thames and Hudson.
Miller, W.J., and P. Capy. 2004. Mobile Genetic Elements as Natural Tools for Genome Evolution, in Mobile Genetic Elements: Protocols and Genomic Applications. Humana Press Inc., Totowa, NJ.
Muthukumar, B., and J.L. Bennetzen. 2004. Isolation and characterization of genomic and transcribed retrotransposon sequences from sorghum. Molecular Genetics and Genomics 271:308-316.
Mooers, A.O., and E.C. Holmes. 2000. The evolution of base composition and phylogenetic inference. Trends in Ecology & Evolution 15:365-369.
Morgante, M. 2006. Plant genome organisation and diversity: the year of the junk! Current Opinion in Biotechnology 17:168-173.
Mroczek, R.J., and R.K. Dawe. 2003. Distribution of retroelements in centromeres and neocentromeres of maize. Genetics 165:809-819.
Natali, L., T. Giordani, M. Buti, and A. Cavallini. 2007. Isolation of Ty1-copia putative LTR sequences and their use as a tool to analyse genetic diversity in Olea europaea. Molecular Breeding 19:255-265.
Navarro-Quezada, A., and D.J. Schoen. 2002. Sequence evolution and copy number of Ty1-copia retrotransposons in diverse plant genomes. Proceedings of the National Academy of Sciences of the United States of America 99:268-273.
Neumann, F., C. Herold, B. Hildebrandt, G. Kobbe, M. Aivado, A. Rong, M. Free, R. Rössig, R. Fenk, and P. Schneider. 2003. Quantitative real-time reverse-transcription polymerase chain reaction for diagnosis of BCR-ABL positive leukemias and molecular monitoring following allogeneic stem cell transplantation, pp. 1-10, Vol. 70. John Wiley & Sons.
Neumann, P., D. Pozarkova, and J. Macas. 2003. Highly abundant pea LTR retrotransposon Ogre is constitutively transcribed and partially spliced. Plant Molecular Biology 53:399-410.
Nobel, P.S. 1988. Environmental biology of agaves and cacti. Science 242:948-949.
Nobel, P.S. 1991. Achievable productivities of certain CAM plants - Basis for high values compared with C3 and C4 plants. New Phytologist 119:183-205.
Nobel, P.S., E. Garciamoya, and E. Quero. 1992. High annual productivity of certain agaves and cacti under cultivation. Plant Cell and Environment 15:329-335.
Noma, K., E. Ohtsubo, and H. Ohtsubo. 1999. Non-LTR retrotransposons (LINEs) as ubiquitous components of plant genomes. Molecular and General Genetics 261:71-79.
Okamoto, H., and H. Hirochika. 2001. Silencing of transposable elements in plants. Trends in Plant Science 6:527-534.
Ohshima, K., M. Hamada, Y. Terai, and N. Okada. 1996. The 3'ends of tRNA-derived short interspersed repetitive elements are derived from the 3'ends of long interspersed repetitive elements, pp. 3756, Vol. 16. Am Soc Microbiol.
Paterson, A.H., J.E. Bowers, and B.A. Chapman. 2004. Ancient polyploidization predating divergence of the cereals, and its consequences for comparative genomics. Proceedings of the National Academy of Sciences of the United States of America 101:9903-9908.
Petrov, D.A. 2002. Mutational equilibrium model of genome size evolution, pp. 531-544, Vol. 61. Elsevier.
Petrov, D.A., and D.L. Hartl. 1999. Patterns of nucleotide substitution in Drosophila and mammalian genomes, pp. 1475, Vol. 96. National Acad Sciences.

Petrov, D.A., T.A. Sangster, J.S. Johnston, D.L. Hartl, and K.L. Shaw. 2000. Evidence for DNA loss as a determinant of genome size, pp. 1060, Vol. 287. AAAS.

Phillips, R.L., S.M. Kaeppler, and P. Olhoft. 1994. Genetic instability of plant tissue cultures: breakdown of normal controls, pp. 5222, Vol. 91. National Acad Sciences.

Pritham, E.J., T. Putliwala, and C. Feschotte. 2007. Mavericks, a novel class of giant transposable elements widespread in eukaryotes and related to DNA viruses. Gene 390:3-17.

Palomino, G., J. Dolezel, I. Mendez, and A. Rubluo. 2003. Nuclear genome size analysis of *Agave tequilana* Weber. Caryologia 56:37-46.

Pearce, S.R., A. Kumar, and A.J. Flavell. 1996a. Activation of the Ty1-copia group retrotransposons of potato (Solanum tuberosum) during protoplast isolation. Plant Cell Reports 15:949-953.

Pearce, S.R., M. Knox, T.H.N. Ellis, A.J. Flavell, and A. Kumar. 2000. Pea Ty1-copia group retrotransposons: transpositional activity and use as markers to study genetic diversity in Pisum. Molecular and General Genetics 263:898-907.

Pearce, S.R., G. Harrison, D.T. Li, J.S. HeslopHarrison, A. Kumar, and A.J. Flavell. 1996b. The Ty1-copia group retrotransposons in Vicia species: Copy number, sequence heterogeneity and chromosomal localisation. Molecular & General Genetics 250:305-315.

Pearce, S.R., C. Stuart-Rogers, M.R. Knox, A. Kumar, T.H.N. Ellis, and A.J. Flavell. 1999. Rapid isolation of plant Ty1-copia group retrotransposon LTR sequences for molecular marker studies. Plant Journal 19:711-717.

Pearce, S.R., G. Harrison, P.J.S. HeslopHarrison, A.J. Flavell, and A. Kumar. 1997. Characterization and genomic organization of Ty1-copia group retrotransposons in rye (Secale cereale). Genome 40:617-625.

Pearce, S.R., U. Pich, G. Harrison, A.J. Flavell, J.S.P. HeslopHarrison, I. Schubert, and A. Kumar. 1996. The Ty1-copia group retrotransposons of Allium cepa are distributed throughout the chromosomes but are enriched in the terminal heterochromatin. Chromosome Research 4:357-364.

Pereira, H.S., A. Barao, M. Delgado, L. Morais-Cecilio, and W. Viegas. 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.

Pelissier, T., S. Tutois, J.M. Deragon, S. Tourmente, S. Genestier, and G. Picard. 1995. Athila, a new retroelement from *Arabidopsis* thaliana. Plant Molecular Biology 29:441-452.

Peterson-Burch, B.D., and D.F. Voytas. 2002. Genes of the Pseudoviridae (Ty1/copia retrotransposons). Molecular Biology and Evolution 19:1832-1845.

Piegu, B., R. Guyot, N. Picault, A. Roulin, A. Saniyal, H. Kim, K. Collura, D.S. Brar, S. Jackson, R.A. Wing, and O. Panaud. 2006. Doubling genome size without polyploidization: Dynamics of retrotransposition-driven genomic expansions in Oryza australiensis, a wild relative of rice. Genome Research 16:1262-1269.

Pourtau, N., B. Lauga, C. Audeon, M.A. Grandbastien, P. Goulas, and J.C. Salvado. 2003. The promoter of the Tnt1A retrotransposon is activated by ozone air pollution in tomato, but not in its natural host tobacco. Plant Science 165:983-992.

Pouteau, S., M.A. Grandbastien, and M. Boccara. 1994. Microbial Elicitors of Plant Defense Responses Activate Transcription of a Retrotransposon. Plant Journal 5:535-542.

Pouteau, S., E. Huttner, M.A. Grandbastien, and M. Caboche. 1991. Specific Expression of the Tobacco Tnt1 Retrotransposon in Protoplasts. Embo Journal 10:1911-1918.

Price, Z., F. Dumortier, D.W. MacDonald, and S. Mayes. 2002. Characterisation of copia-like retrotransposons in oil palm (Elaeis guineensis Jacq.). Theoretical and Applied Genetics 104:860-867.

Presting, G.G., L. Malysheva, J. Fuchs, and I.Z. Schubert. 1998. A Ty3/Gypsy retrotransposon-like sequence localizes to the centromeric regions of cereal chromosomes. Plant Journal 16:721-728.

Ramallo, E., R. Kalendar, A.H. Schulman, and J.A. Martinez-Izquierdo. 2008. Reme1, a Copia retrotransposon in melon, is transcriptionally induced by UV light. Plant Molecular Biology 66:137-150.

Rico-Cabanas, L., and J.A. Martinez-Izquierdo. 2007. CIRE1, a novel transcriptionally active Ty1-copia retrotransposon from Citrus sinensis. Molecular Genetics and Genomics 277:365-377.

Ritsema, T., and S. Smeekens. 2003. Fructans: beneficial for plants and humans. Current Opinion in Plant Biology 6:223-230.

Rubin, G.M., M.G. Kidwell, and P.M. Bingham. 1982. The molecular basis of PM hybrid dysgenesis: the nature of induced mutations, pp. 987-994, Vol. 29. Elsevier.

SanMiguel, P., A. Tikhonov, Y.K. Jin, N. Motchoulskaia, D. Zakharov, A. MelakeBerhan, P.S. Springer, K.J. Edwards, M. Lee, Z. Avramova, and J.L. Bennetzen. 1996. Nested retrotransposons in the intergenic regions of the maize genome. Science 274:765-768.

Sasaki, T., T. Matsumoto, K. Yamamoto, K. Sakata, T. Baba, Y. Katayose, J.Z. Wu, Y. Niimura, Z.K. Cheng, Y. Nagamura, B.A. Antonio, H. Kanamori, S. Hosokawa, M. Masukawa, K. Arikawa, Y. Chiden, M. Hayashi, M. Okamoto, T. Ando, H. Aoki, K. Arita, M. Hamada, C. Harada, S. Hijishita, M. Honda, Y. Ichikawa, A. Idonuma, M. Iijima, M. Ikeno, S. Ito, T. Ito, Y. Ito, Y. Ito, A. Iwabuchi, K. Kamiya, W. Karasawa, S. Katagiri, A. Kikuta, N. Kobayashi, I. Kono, K. Machita, T. Maehara, H. Mizuno, T. Mizubayashi, Y. Mukai, H. Nagasaki, M. Nakashima, Y. Nakama, Y. Nakamichi, M. Nakamura, N. Namiki, M. Negishi, I. Ohta, N. Ono, S. Saji, K. Sakai, M. Shibata, T. Shimokawa, A. Shomura, J.Y. Song, Y. Takazaki, K. Terasawa, K. Tsuji, K. Waki, H. Yamagata, H. Yamane, S. Yoshiki, R. Yoshihara, K. Yukawa, H.S. Zhong, H. Iwama, T. Endo, H. Ito, J.H. Hahn, H.I. Kim, M.Y. Eun, M. Yano, J.M. Jiang, and T. Gojohori. 2002. The genome sequence and structure of rice chromosome 1. Nature 420:312-316.

Schulman, A.H., A.J. Flavell, and T.H.N. Ellis. 2004. The application of LTR retrotransposons as molecular markers in plants, pp. 145-174, Vol. 260. Springer.

Sheldon, S. 1980. Ethnobotany of agave-lecheguilla and yucca-carnerosana in mexico zona ixtlera. Economic Botany 34:376-390.

Smarda, P., P. Bures, L. Horova, and O. Rotreklova. 2008. Intrapopulation genome size dynamics in Festuca pallens. Annals of Botany 102:599-607.

Sabot, F., and A.H. Schulman. 2006. Parasitism and the retrotransposon life cycle in plants: a hitchhiker's guide to the genome. Heredity 97:381-388.

Sabot, F., R. Guyot, T. Wicker, N. Chantret, B. Laubin, B. Chalhoub, P. Leroy, P. Sourdille, and M. Bernard. 2005. Updating of transposable element annotations from large wheat genomic sequences reveals diverse activities and gene associations. Molecular Genetics and Genomics 274:119-130.

Saitou, N., and M. Nei. 1987. The Neighbor-Joining Method - a New Method for Reconstructing Phylogenetic Trees. Molecular Biology and Evolution 4:406-425.

Salazar, M., E. Gonzalez, J.A. Casaretto, J.M. Casacuberta, and S. Ruiz-Lara. 2007. The promoter of the TLC1.1 retrotransposon from Solanum chilense is activated by multiple stress-related signaling molecules. Plant Cell Reports 26:1861-1868.

Saldana, O.I. 2006. Carbohydrate metabolism in the leaf of the *Agave tequilana* (Weber) plant. PhD thesis, University of Sussex, UK.

Sandmeyer, S.B., L.J. Hansen, and D.L. Chalker. 1990. Integration Specificity of Retrotransposons and Retroviruses. Annual Review of Genetics 24:491-518.
SanMiguel, P., and J.L. Bennetzen. 1998. Evidence that a recent increase in maize genome size was caused by the massive amplification of intergene retrotransposons. Annals of Botany 82:37-44.
SanMiguel, P., B.S. Gaut, A. Tikhonov, Y. Nakajima, and J.L. Bennetzen. 1998. The paleontology of intergene retrotransposons of maize. Nature Genetics 20:43-45.
SanMiguel, P., A. Tikhonov, Y.K. Jin, N. Motchoulskaia, D. Zakharov, A. MelakeBerhan, P.S. Springer, K.J. Edwards, M. Lee, Z. Avramova, and J.L. Bennetzen. 1996. Nested retrotransposons in the intergenic regions of the maize genome. Science 274:765-768.
SCFI. 1997. Norma Oficial Mexicana. NOM-006-SCFI-1994. Bebidas Alcoholicas-Tequila-Especificaciones. Diario Oficial de La Federacion. Ciudad de Mexico.
Schmidt, T. 1999. LINEs, SINEs and repetitive DNA: non-LTR retrotransposons in plant genomes. Plant Molecular Biology 40:903-910.
Shedlock, A.M., and N. Okada. 2000. SINE insertions: powerful tools for molecular systematics. Bioessays 22:148-160.
Shirasu, K., A.H. Schulman, T. Lahaye, and P. Schulze-Lefert. 2000. A contiguous 66-kb barley DNA sequence provides evidence for reversible genome expansion. Genome Research 10:908-915.
Slauson, L.A. 2001. Insights on the pollination biology of Agave (Agavaceae). Haseltonia:10-23.
Smith, J.A.C., and P.S. Nobel. 1986. Water movement and storage in a desert succulant- anatomy and rehydration kinetics for leaves of *Agave deserti*. Journal of Experimental Botany 37:1044-1053.
Smykal, P. 2006. Development of an efficient retrotransposon-based fingerprinting method for rapid pea variety identification. Journal of Applied Genetics 47:221-230.
Stuart-Rogers, C., and A.J. Flavell. 2001. The evolution of Ty1-copia group retrotransposons in gymnosperms. Molecular Biology and Evolution 18:155-163.
Sugimoto, K., S. Takeda, and H. Hirochika. 2000. MYB-related transcription factor NtMYB2 induced by wounding and elicitors is a regulator of the tobacco retrotransposon Tto1 and defense-related genes. Plant Cell 12:2511-2527.
Sun, F.J., S. Fleurdepine, C. Bousquet-Antonelli, G. Caetano-Anolles, and J.M. Deragon. 2007. Common evolutionary trends for SINE RNA structures. Trends in Genetics 23:26-33.
Suoniemi, A., A. Narvanto, and A.H. Schulman. 1996a. The BARE-1 retrotransposon is transcribed in barley from an LTR promoter active in transient assays. Plant Molecular Biology 31:295-306.
Suoniemi, A., K. AnamthawatJonsson, T. Arna, and A.H. Schulman. 1996b. Retrotransposon BARE-1 is a major, dispersed component of the barley (Hordeum vulgare L) genome. Plant Molecular Biology 30:1321-1329.
Syed, N.H., S. Sureshsundar, M.J. Wilkinson, B.S. Bhau, J.J.V. Cavalcanti, and A.J. Flavell. 2005. Ty1-copia retrotransposon-based SSAP marker development in cashew (Anacardium occidentale L.). Theoretical and Applied Genetics 110:1195-1202.
Tahara, M., T. Aoki, S. Suzuka, H. Yamashita, M. Tanaka, S. Matsunaga, and S. Kokumai. 2004. Isolation of an active element from a high-copy-number family of retrotransposons in the sweetpotato genome. Molecular Genetics and Genomics 272:116-127.
Takata, M., A. Kiyohara, A. Takasu, Y. Kishima, H. Ohtsubo, and Y. Sano. 2007. Rice transposable elements are characterized by various methylation environments in the genome. Bmc Genomics 8.
Takeda, S., K. Sugimoto, H. Otsuki, and H. Hirochika. 1998. Transcriptional activation of the tobacco retrotransposon Tto1 by wounding and methyl jasmonate. Plant Molecular Biology 36:365-376.

Takeda, S., K. Sugimoto, H. Otsuki, and H. Hirochika. 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. Plant Journal 18:383-393.

Takano, M., H. Kanegae, T. Shinomura, A. Miyao, H. Hirochika, and M. Furuya. 2001. Isolation and characterization of rice phytochrome A mutants, pp. 521, Vol. 13. Am Soc Plant Biol.

Timmerman-Vaughan, G.M., M. Pither-Joyce, and G. Clover. 2005. A Positive-Control PCR Assay That Is Based on Chloroplast ndhB Sequences and Applicable to Nucleic Acid Extractions From Diverse Plant Taxa, pp. 77-78, Vol. 23. New York: Published by Mary Ann Liebert, Inc., Publishers for the International Society for Plant Molecular Biology, c1983-.

Todorovska, E. 2007. Retrotransposons and their role in plant - Genome evolution. Biotechnology & Biotechnological Equipment 21:294-305.

Tam, S.M., C. Mhiri, A. Vogelaar, M. Kerkveld, S.R. Pearce, and M.A. Grandbastien. 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.

Tamura, K., J. Dudley, M. Nei, and S. Kumar. 2007. MEGA4: Molecular evolutionary genetics analysis (MEGA) software version 4.0. Molecular Biology and Evolution 24:1596-1599.

Tao, N.G., J. Xu, Y.J. Cheng, L. Hong, W.W. Guo, H.L. Yi, and X.X. Deng. 2005. Isolation and characterization of Copia-like retrotransposons from 12 sweet orange (Citrus sinensis Osbeck) cultivars. Journal of Integrative Plant Biology 47:1507-1515.

Teo, C.H., S.H. Tan, C.L. Ho, Q.Z. Faridah, Y.R. Othman, J.S. Heslop-Harrison, R. Kalendar, and A.H. Schulman. 2005. Genome constitution and classification using retrotransposon-based markers in the orphan crop banana. Journal of Plant Biology 48:96-105.

Thomas, C.A. 1971. GENETIC ORGANIZATION OF CHROMOSOMES. Annual Review of Genetics 5:237-&.

Tissue, D.T., and P.S. Nobel. 1988. Parent ramet connections in *Agave deserti*- influences of carbohydrates on growth. Oecologia 75:266-271.

Valenzuela-Zapata, A.G. 1997. El Agave Tequilero, su cultivo e industria. Lıtteris Editores, Mexico, 1997, pp. 1–204.

Valenzuela-Zapata, A.G., and G.P. Nabhan. 2003. Tequila. A natural and cultural history. The University of Arizona Press, Tuscon, Arizona.

Vega, K.G., C. Diaz, A.N. Cedillo, and J. Simpson. 2006. AFLP analysis of *Agave tequilana* varieties. Plant Science 170 pp. 904-909.

Vega, K.G., M.G. Chavira, O.M. de la Vega, J. Simpson, and G. Vandemark. 2001. Analysis of genetic diversity in *Agave tequilana* var. Azul using RAPD markers. Euphytica 119:335-341.

Vance, V., and H. Vaucheret. 2001. RNA silencing in plants--defense and counterdefense, pp. 2277, Vol. 292. AAAS.

Varagona, M.J., M. Purugganan, and S.R. Wessler. 1992. Alternative splicing induced by insertion of retrotransposons into the maize waxy gene, pp. 811, Vol. 4. Am Soc Plant Biol.

Verastegui, M.A., C.A. Sanchez, N.L. Heredia, and J.S. GarciaAlvarado. 1996. Antimicrobial activity of extracts of three major plants from the Chihuahuan desert. Journal of Ethnopharmacology 52:175-177.

Vereyken, I.J., V. Chupin, F.A. Hoekstra, S.C.M. Smeekens, and B. de Kruijff. 2003. The effect of fructan on membrane lipid organization and dynamics in the dry state. Biophysical Journal 84:3759-3766.

Vernhettes, S., M.A. Grandbastien, and J.M. Casacuberta. 1997. In vivo characterization of transcriptional regulatory sequences involved in the defence-associated expression of the tobacco retrotransposon Tnt1. Plant Molecular Biology 35:673-679.
Vernhettes, S., M.A. Grandbastien, and J.M. Casacuberta. 1998. The evolutionary analysis of the Tnt1 retrotransposon in Nicotiana species reveals the high variability of its regulatory sequences. Molecular Biology and Evolution 15:827-836.
Vaucheret, H., and M. Fagard. 2001. Transcriptional gene silencing in plants: targets, inducers and regulators, pp. 29-35, Vol. 17. Elsevier.
Vignols, F., J. Rigau, M.A. Torres, M. Capellades, and P. Puigdomenech. 1995. The brown midrib3 (bm3) mutation in maize occurs in the gene encoding caffeic acid O-methyltransferase, pp. 407, Vol. 7. Am Soc Plant Biol.
Verries, C., C. Bes, P. This, and C. Tesniere. 2000. Cloning and characterization of Vine-1, a LTR-retrotransposon-like element in Vitis vinifera L., and other Vitis species. Genome 43:366-376.
Vershinin, A.V., T.R. Allnutt, M.R. Knox, M.J. Ambrose, and T.H.N. Ellis. 2003. Transposable elements reveal the impact of introgression, rather than transposition, in Pisum diversity, evolution, and domestication. Molecular Biology and Evolution 20:2067-2075.
Vicient, C.M., and A.H. Schulman. 2002. Copia-like retrotransposons in the rice genome: few and assorted. Genome lett. 1, 35-47.
Vicient, C.M., A. Suoniemi, K. Anamthamat-Jonsson, J. Tanskanen, A. Beharav, E. Nevo, and A.H. Schulman. 1999. Retrotransposon BARE-1 and its role in genome evolution in the genus Hordeum. Plant Cell 11:1769-1784.
Vicient, C.M., M.J. Jaaskelainen, R. Kalendar, and A.H. Schulman. 2001. Active retrotransposons are a common feature of grass genomes. Plant Physiology 125:1283-1292.
Vicient, C.M., R. Kalendar, K. Anamthawat-Jonsson, A. Suoniemi, and A.H. Schulman. 1999a. Structure, functionality, and evolution of the BARE-1 retrotransposon of barley. Genetica 107:53-63.
Vieira, M.C., T. Heinze, R. Antonio-Cruz, and A.M. Mendoza-Martinez. 2002. Cellulose derivatives from cellulosic material isolated from Agave lechuguilla and fourcroydes. Cellulose 9:203-212.
Vinogradov, A.E. 2003. Selfish DNA is maladaptive: evidence from the plant Red List. Trends in Genetics 19:609-614.
Vitte, C., and O. Panaud. 2003. Formation of solo-LTRs through unequal homologous recombination counterbalances amplifications of LTR retrotransposons in rice Oryza sativa L. Molecular Biology and Evolution 20:528-540.
Vitte, C., and O. Panaud. 2005. LTR retrotransposons and flowering plant genome size: emergence of the increase/decrease model. Cytogenetic and Genome Research 110:91-107.
Vitte, C., and J.L. Bennetzen. 2006. Analysis of retrotransposon structural diversity uncovers properties and propensities in angiosperm genome evolution. Proceedings of the National Academy of Sciences of the United States of America 103:17638-17643.
Vos, P., R. Hogers, M. Bleeker, M. Reijans, T. Vandelee, M. Hornes, A. Frijters, J. Pot, J. Peleman, M. Kuiper, and M. Zabeau. 1995. Aflp - a New Technique for DNA-Fingerprinting. Nucleic Acids Research 23:4407-4414.
Voytas, D.F., and F.M. Ausubel. 1988. A Copia-Like Transposable Element Family in *Arabidopsis*-Thaliana. Nature 336:242-244.
Voytas, D.F., M.P. Cummings, A. Konieczny, F.M. Ausubel, and S.R. Rodermel. 1992. Copia-Like Retrotransposons Are Ubiquitous among Plants. Proceedings of the National Academy of Sciences of the United States of America 89:7124-7128.

Wang, S.P., N. Liu, K.M. Peng, and Q.F. Zhang. 1999. The distribution and copy number of copia-like retrotransposons in rice (Oryza sativa L.) and their implications in the organization and evolution of the rice genome. Proceedings of the National Academy of Sciences of the United States of America 96:6824-6828.

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

Wessler, S.R. 1996. Plant retrotransposons: Turned on by stress. Current Biology 6:959-961.

Wessler, S.R. 2001. Plant transposable elements. A hard act to follow. Plant Physiology 125:149-151.

Wessler, S.R. 2006. Transposable elements and the evolution of eukaryotic genomes. Proceedings of the National Academy of Sciences of the United States of America 103:17600-17601.

Wessler, S.R., T.E. Bureau, and S.E. White. 1995. Ltr-Retrotransposons and Mites - Important Players in the Evolution of Plant Genomes. Current Opinion in Genetics & Development 5:814-821.

Wessler, S.R., G. Baran, and M. Varagona. 1987. The Maize Transposable Element Ds Is Spliced from Rna. Science 237:916-918.

White, S.E., L.F. Habera, and S.R. Wessler. 1994. Retrotransposons in the Flanking Regions of Normal Plant Genes - a Role for Copia-Like Elements in the Evolution of Gene Structure and Expression. Proceedings of the National Academy of Sciences of the United States of America 91:11792-11796.

Wicker, T., and B. Keller. 2007. Genome-wide comparative analysis of copia retrotransposons in Triticeae, rice, and *Arabidopsis* reveals conserved ancient evolutionary lineages and distinct dynamics of individual copia families. Genome Research 17:1072-1081.

Wicker, T., N. Stein, L. Albar, C. Feuillet, E. Schlagenhauf, and B. Keller. 2001. Analysis of a contiguous 211 kb sequence in diploid wheat (Triticum monococcum L.) reveals multiple mechanisms of genome evolution. Plant Journal 26:307-316.

Wicker, T., R. Guyot, N. Yahiaoui, and B. Keller. 2003a. CACTA transposons in Triticeae. A diverse family of high-copy repetitive elements, pp. 52, Vol. 132. Am Soc Plant Biol.

Wicker, T., N. Yahiaoui, R. Guyot, E. Schlagenhauf, Z.D. Liu, J. Dubcovsky, and B. Keller. 2003b. Rapid genome divergence at orthologous low molecular weight glutenin loci of the A and Am genomes of wheat, pp. 1186, Vol. 15. Am Soc Plant Biol.

Wikstrom, N., V. Savolainen, and M.W. Chase. 2001. Evolution of the angiosperms: calibrating the family tree. Proceedings of the Royal Society of London Series B-Biological Sciences 268:2211-2220.

Wilhelm, M., and F.X. Wilhelm. 2001. Reverse transcription of retroviruses and LTR retrotransposons. Cellular and Molecular Life Sciences 58:1246-1262.

Winter, K., and J.A.C. Smith. 1996. An introduction to crassulacean acid metabolism. Biochemical principles and ecological diversity, p. 1-13, *In* K. Winter and J. A. C. Smith, eds. Crassulacean Acid Metabolism - Biochemistry, Ecophysiology and Evolution, Vol. 114.

Xiao, W.M., W. Sakamoto, and Sodmergen. 2004. Isolation and characterization of Ty1/copia-like reverse transcriptase sequences from mung bean. Acta Botanica Sinica 46:582-587.

Xiong, Y., and T.H. Eickbush. 1990. Origin and Evolution of Retroelements Based Upon Their Reverse-Transcriptase Sequences. Embo Journal 9:3353-3362.

Yang, N., and H.H. Kazazian. 2006. L1 retrotransposition is suppressed by endogenously encoded small interfering RNAs in human cultured cells. Nature Structural & Molecular Biology 13:763-771.

Yu, J., J. Wang, W. Lin, and S.G. Li. 2005. The Genomes of Oryza sativa: A history of duplications. Plos Biology 3:266-281.
Zhang, X.Y., and S.R. Wessler. 2005. BoS: A large and diverse family of short interspersed elements (SINEs) in Brassica oleracea. Journal of Molecular Evolution 60:677-687.
Zuckerkandl, E., and L. Pauling. 1965. Evolutionary divergence and convergence in proteins. Evolving Genes and Proteins, edited by V. Bryson and H.J. Vogel. Academic Press, New York.:pp. 97-166