UNIVERSITY OF CALIFORNIA
RIVERSIDE

A Molecular Genetic Approach to Evaluate a Novel Seedless Phenotype Found in Tango, a New Variety of Mandarin Developed From Gamma-Irradiated W. Murcott

Doctor of Philosophy

in

Genetics, Genomics and Bioinformatics

by

Jennifer Robyn Crowley

December 2011

Dissertation Committee:

Dr. Mikeal Roose, Chairperson
Dr. Linda Walling
Dr. Katherine Borkovich

The Dissertation of Jennifer Robyn Crowley is approved:

_____

_____

Committee Chairperson

University of California, Riverside

Acknowledgments

I would like to acknowledge and give my many thanks to several people who contributed both their time and their effort towards my dissertation work. I would like to thank Dr. David Carter for his instruction regarding the correct use of the Leica SP2 confocal microscope, as well as for his help in gathering quality computerized images. I would like to thank Dr. Chandrika Ramadugu and Dr. Manjunath Keremane for their assistance with the use of the Stratagene MX thermocycler, as well as for their help in qPCR data interpretation. To Dr. Chandrika Ramadugu, I would also like to extend my gratitude for the contribution she made towards my research project by providing me with clones of a DNA sequence of interest. I would like to thank Dr. Barbara Jablonska for her help in the radiolabeling process for Southern blotting. I would like to thank Dr. Lisa Mu for all of her data related to the methylation study of W. Murcott and Tango, which added greatly to the epigenetic analysis of these two mandarin varieties. I extend many thanks to Dr. Claire Federici for ensuring that all necessary reagents and materials related to my project were ordered and kept on file for my convenience. Lastly, I would like to thank Dr. Mikeal Roose for his support and guidance throughout this dissertation project.

ABSTRACT OF THE DISSERTATION

A Molecular Genetic Approach to Evaluate a Novel Seedless Phenotype Found in Tango, a New Variety of Mandarin Developed From Gamma-Irradiated W. Murcott

by

Jennifer Robyn Crowley

Doctor of Philosophy, Graduate Program in Genetics, Genomics and Bioinformatics
University of California, Riverside, December 2011
Dr. Mikeal Roose, Chairperson

Since the beginning of its first known cultivation as described in ancient European literature (circa 300 BCE) citrus fruit has been one of the most important consumer products in the world. Currently, there is a great deal of research in developing new citrus varieties and in improving existing citrus varieties. Creating and improving citrus varieties can be achieved in one of several different ways: cross-breeding, genetic transformation, artificial selection, and irradiation. Molecular alterations within plant tissue exposed to radiation, such as genetic mutations, can and do occur, making the method of plant irradiation a desirable choice when attempting to create unique citrus varieties.

The intent of this study is to evaluate and compare the molecular genetics of one particular variety of citrus, a seedless mandarin-producing Tango, with its predecessor variety, a seedy mandarin-producing W. Murcott. The Tango variety was developed and

propagated from gamma-irradiated W. Murcott budwood. In the case of Tango, fruit produced by a mature tree retained the same characteristics of size, color, easiness to peel, and sweetness as that produced by W. Murcott; Tango fruit differed from W. Murcott fruit only in its seedlessness.

To determine the underlying molecular differences between W. Murcott and Tango, molecular regions prone to alteration due to radiation exposure were targeted for analysis. These included chromosomes, LTR (long-terminal repeat) retrotransposon regions, and methylated/unmethylated CpG regions. For chromosome analysis, stained images of chromosome spreads taken from W. Murcott and Tango cells frozen in their meiotic metaphase state were captured via a Leica SP2 Confocal microscope. Images were compared and contrasted between the two citrus varieties to determine any significant differences in chromosome alignment. CIRE1, a sequenced, intact LTR retrotransposon in the *Citrus sinensis* genome was targeted for quantitative analysis in the W. Murcott and Tango genomes. Primer sets were designed based on the CIRE1 sequence found within the NCBI database archives, targeting three regions of CIRE1: a long-terminal repeat, an integrase gene, and a reverse transcriptase gene. CpG region methylation patterns were compared and contrasted between W. Murcott and Tango DNA extractions using LUMA analysis followed by pyrosequencing. *HpaII* (methylation insensitive) and *MseI* (methylation sensitive) restriction enzymes were used to digest W. Murcott and Tango DNA subsequent to pyrosequencing; the digestions were compared with their respective sequenced untreated controls as well as to each other to determine any significant deviations in CpG island methylation levels. Methylation-sensitive Amplified

Fragment Length Polymorphism (ms-AFLP) analysis, which targeted methylated cytosine sites in the *Citrus* genome, was also performed by Lisa Mu to determine whether or not polymorphic bands between W. Murcott and Tango could be visualized on a polyacrylamide gel.

The results from experimentation showed that W. Murcott and Tango are nearly genetically identical. Both W. Murcott and Tango were found to have 14 copies of the CIRE1 retrotransposon in their genomes, and there were no significant differences in the percentage of methylated cytosines within their genomes (ranging from 53.1% to 58.6%) nor were there deviations in methylated cytosine patterns. However, there are some notable molecular genetic differences between the two varieties; in particular, chromosome misalignment in the meiotic metaphase state was detected in several images of Tango chromosome spreads, while only normal chromosome alignment in the meiotic metaphase state was ever detected in the images taken of W. Murcott chromosome spreads. This suggests that chromosomal abnormalities resulting from gamma radiation exposure could be the main underlying genetic cause of seedless fruit produced by Tango citrus plants.

Table of Contents

Introduction to Dissertation

History of Citrus

Citrus are among the most economically and commercially important fruit crops in the world; their consumption has increased steadily with an increase in population. Citrus is believed to have originated in Southeast Asia. Chief areas where commercial citrus is grown include southern China, the Mediterranean Basin (as well as southern Spain), South Africa, Australia, the southernmost United States and parts of South America (Webber *et al* 1967). It is now known that the sweet orange was cultivated for many centuries in China and appears to have attained a complex phase of cultivation before being introduced to Europeans (Webber *et al* 1967). *Monograph on the Oranges of Wên-chou, Chekiang,* dated from 1178 AD, is the oldest surviving book on the orange; one can presume that oranges have been an often-discussed topic in Chinese literature, dating back to the time of Ch'u Yuan, who had written a poem entitled "Li Sao" or "Falling into Trouble," in which many plants, trees, and fruits from that time period are described (314 B.C.) (Webber *et al* 1967). Most sweet orange seedlings develop into trees bearing sweet, edible fruit, even in the absence of improved varieties (Webber *et al* 1967). In all likelihood, citrus fruits that first caught the attention of ancient civilizations had already undergone genetic development through the processes of natural evolution (Webber *et al* 1967). Due to the pleasing characteristics of the fruit as a food source, citrus fruits were

selected as fruits worthy of cultivation, and in all probability artificial selection for the best individuals for propagation had been ongoing for centuries before such individuals were introduced for cultivation in European countries (Webber *et al* 1967).

Development of the Citrus Industry in California

While the cultivation of citrus in California has been ongoing since the days of the Spanish missions, it "remained a minor crop until sometime in the late nineteenth century" (Lawton *et al* 1989). Lawton et al describes a brief history of the rise of the citrus industry in California: in 1870, there existed a mere 8,000 orange trees in Los Angeles County with fewer than 35,000 trees in the entire state of California. Towards the end of the 1860s in southern California, cattle ranch owners were forced to sell their Mexican land grants that had been subjected to years of drought; these lands were later purchased by increasing numbers of Americans migrating from the East. A succession of colony towns were developed east of Los Angeles over the following two decades. It was these irrigated settlements in desert-like surroundings that formed the groundwork of a vast citrus empire across inland southern California. The "Citrus Belt," as it was eventually called, extended over 70 miles alongside plateaus "bordering the San Gabriel and San Bernardino mountain ranges; from Pasadena in the west to Redlands in the east". By 1880, over one million orange trees were growing in this area, and thousands more were planted each year.

The Taxonomy of Citrus

The nomenclature and systematics of *Citrus* are complicated and the complete number of natural species is uncertain; the majority of named species are clonally propagated hybrids, and there exists genetic proof that certain wild, true-breeding species are of hybrid origin (Gmitter *et al* 1990). The actual number of species belonging to *Citrus* as well as their genetic ties with one another are sources for debate in *Citrus* taxonomy (Pang *et al* 2006). The general consensus, based on the studies of Scora (1975) as well as Barrett and Rhodes (1976), is that cultivated *Citrus* is derived of the three following species: *C. medica* L., *C. reticulata* Blanco, and *C. maxima* (Burm.) Merrill. *C. halimii* B.C. Stone (mountain citron) may represent a fourth ancestral species but it has apparently not contributed genes to the cultivated groups. *Citrus* varieties such as limes, grapefruit, and lemons are considered genetic offshoots of these three true species; either through cross-hybridization amongst the true species or through cross-hybridization between them and closely related species (Pang *et al* 2006). Like most other fruits, citrus varieties can be traced back to a single parent tree or individual mutant branch which has been propagated by vegetative or asexual means, such as cuttage or graftage (Gmitter *et al* 1990). A more precise term was assigned to such plants by Webber (1903) who proposed the term "clone" (Webber *et al* 1967). Citrus cultivars are therefore clonal varieties.

Citrus Crop Improvement

With the rising consumer population, the demand for more citrus fruit having desirable characteristics is steadily increasing. In the *Citrus* groups, the development and

implementation of genomic tools could prove vital in further crop improvements (Terol *et al* 2008). The most up-to-date genomic technologies can be used effectively in citrus crop improvement studies, with much potential for genomic research to enhance the field of molecular biology in regards to citrus crop improvement (Sonah *et al* 2011). One of the main goals for the improvement of fresh fruit in the world has included the development of crops yielding fruits that are seedless, easy-peeling, flavorful and aromatic, and carry a long ripening period. Because of the location of the seeds, seedlessness is a very desirable fruit trait in citrus grown for fresh consumption from the viewpoint of the consumer. Due to their commercial value, horticulturists and breeders have long been interested in the development of seedless cultivars (Raza *et al* 2003). Breeders have explored various methods by which progeny bearing nearly seedless fruit can be developed. Strategies have included "conventional and nonconventional breeding techniques" (Raza *et al* 2003) such as cross hybridization, protoplast fusion, genetic transformation, and irradiation. Conventional breeding methods can be laborious, time-consuming and expensive; therefore, mutation breeding has been turned to as an alternative way of achieving new citrus cultivars. Techniques that have the ability to generate mutations, such as radiation or chemical mutagens, are advantageous for increasing variability in crop species because mutations are spontaneous and will oftentimes occur with an extremely low frequency (Sutarto *et al* 2009). Other methods, such as heat or temperature shocks, seed centrifugation, and ageing of pollen grains and seeds have also been attempted in order to induce mutations in plants, but these methods have been proven to be less effective (Van Harten 1998). Mutation-inducing methods have contributed substantially to plant

improvement throughout the world, and they have made a strong influence on the productivity and economic value of some crops (Sutarto *et al* 2009). The method of utilizing radiation as a way to produce favorable effects in crops is well established (Anonymous 1960). For example, radioisotopes have been widely used in studies involving plant nutrition and research has been performed using radiation sources to preserve selected types of agricultural produce (Anonymous 1960). One of the most important advantages that can come from utilizing radiation on crops is an increase in crop yields by creating new varieties of plants through radiation-induced genetic changes (Anonymous 1960). High energy radiation is often selected for inducing mutations in plants, with X-rays and gamma rays most often used due to their high energy levels and lack of residual radioactivity in the irradiated plant material (Van Harten 1998). It is fortunate that genetic changes brought about through radiation exposure, while generally deleterious in organisms such as humans, can be withstood and oftentimes survived by simpler forms of life such as plants (Anonymous 1960). Mutation techniques have significantly contributed to plant improvement worldwide, and have made an outstanding impact on the productivity and economic value of some crops (Sutarto *et al* 2009). For instance, in India, successful plant varieties brought about through mutation breeding have included cereals, grain legumes,oil seeds fiber crops, vegetables and ornamentals; with the mutant varieties deemed superior to their non-mutant counterparts (Chopra 2005). Inducing mutations by Gamma-rays has been effectively used with several species of *Citrus* (Raza *et al* 2003, Huang *et al* 2003, Sutarto *et al* 2009). One example of such mutation breeding is provided by W. Murcott, a diploid mandarin cultivar that was

treated with Gamma-rays resulting in a low-seeded mutant, later released as 'Tango' (Williams and Roose 2010).

Development of 'Tango' Mandarin by Irradiation of W. Murcott Budwood

The W. Murcott Afourer mandarin has been widely grown in California. This is due to factors such as good fruit quality, non-granulation, easily peeled fruit, and good fruit yields (Wright *et al* 2007). More importantly for the consumer, it has the capability of being seedless. Seedlessness in W. Murcott mandarins does require that the fruit not be cross-pollinated (Wright *et al* 2007). This is due to the fact that W. Murcott is a self-incompatible mandarin, meaning that when self-pollinated, gametes are unable to effect fertilization, thus preventing the formation of seeds (Ollitrault *et al* 2007). One of the primary ways to ensure that there will be no cross-pollination is by planting this variety in isolated areas, away from other citrus varieties; however, it is not always easy to isolate one citrus from others (Khan and Chao 2004). Recently, some California growers cover whole rows of mandarin trees with screen during flowering to prevent cross pollination and the resulting seediness which greatly reduces fruit value. Because ensuring that W. Murcott trees remain in isolation to prevent cross-pollination is time-consuming, laborious, and carries a risk of financial loss should the trees accidentally be cross-pollinated, development of a cultivar that will be seedless even when cross-pollinated was initiated. In citrus breeding, ionizing radiation has been a popular technique to induce mutations in citrus. Studies have shown that, when treated with ionizing radiation, plants subsequently yield very few seeds or fail to produce them

altogether (Viccini *et al* 2002, Raza *et al* 2003, Huang *et al* 2003). Relatively high levels of radiation exposure can produce mutant citrus cultivars that have few seeds per fruit (Hearn 1984, Vardi *et al* 1996, Roose and Williams 2007a). The strategy involves gathering budwood from a citrus tree of interest and exposing the budwood to the desired level of ionizing radiation; it is believed that radiation induces genetic alterations in meristem cells of the budwood. If these cells give rise to a shoot, this variant can later be selected for propagation. There have been instances in which seedless cultivars have been selected from variations in budwood in seedy cultivars (Ye *et al* 2009). W. Murcott budwood was therefore treated with ionizing radiation (namely, γ-rays) in an attempt to induce variations in the budwood which might result in a seedless cultivar. Trees were propagated from the irradiated buds and evaluated for seediness and other traits. The most promising of these selections (known during testing as W. Murcott IR1) has been designated 'Tango' and is similar to W. Murcott in most characters except for very low seed content (0.2 seeds/fruit) and low pollen viability (Williams and Roose 2010). Studies of Tango pollen and pollinations with Tango pollen on Clementine mandarins indicate that Tango pollen has very low germination (<5%) and is smaller than W. Murcott pollen (Roose and Williams 2007b). In hand pollination trials Tango has not set fruit or seed in Clementine mandarins (Roose and Williams 2007b). It is important to know that bud variation stems from a change in the genetic constitution, which may include changes in chromosomal number, chromosomal aberrations, and gene mutations (Ye *et al* 2009). In order to better understand the potential underlying genetic alterations

that result in seedlessness in Tango, various known effects of ionizing radiation will be examined in this study.

Alterations in Chromosomes Exposed to Radiation

Chromosomal rearrangements are one of the most frequently produced classes of mutation that result from the action of both physical and chemical mutagenic agents (Viccini *et al* 2002). In the 1950s, M. S. Swaminathan investigated the effects of chemical and radiation treatments in various types of crop plants, including wheat, barley, and cotton (Natarajan 2005). He found that the treatments induced mutations in the chromosomes of the plants, detected by chromosomal aberrations in the somatic cells of root meristems and in the gametic cells of pollen (Natarajan 2005). Karl Sax was the first to note that the rate of chromosomal mutation increases as the dosage of radiation increases and that X-ray radiation used in his experiments resulted in chromosomal aberrations such as translocations, inversions, and deletions (Sax 1938). These anomalies can alter gene expression if there are changes in the structure of genes, and if the affected cells give rise to gametes, hereditary changes can be expected to follow. Chromosomal abnormalities can also be associated with a reduction in fertility (Viccini *et al* 2002). This is believed to occur because the affected individuals are heterozygous for chromosome rearrangements and consequently have abnormal chromosomal pairings during meiosis. This results in gametes carrying deficiencies and/or duplications, leading to sterility (de

Vozzi *et al* 2009). In such instances, seedlessness is brought about through the joining together of these genetically unbalanced gametes with other gametes, forming a genetically abnormal zygote that is subsequently aborted.

Citrus chromosomes (2n=18) are small (around 2 μm) and have similar morphologies, hindering a detailed cytogenetic characterization by classic methods. Multicolor fluorescent staining and in situ hybridization provides a useful approach to the problem of chromosome analysis and identification (Savage and Simpson 1993, Roose *et al* 1998, Kato *et al* 2004, Findley *et al* 2010). Such technology has been applied in important cereal species and many other species, including those with small chromosomes such as *Arabidopsis thaliana* (Kamm *et al* 1995). According to Carvalho *et al*, the species of *Citrus* are variable in the amount and location of heterochromatic blocks (Carvalho *et al* 2005). Such blocks, or bands, are more easily observed after double staining with the fluorochromes chromomycin A3 (CMA3) and 4,6-diamidino-2-phenylindole (DAPI) (Carvalho *et al* 2005). This is explained by the high affinity that CMA3 has towards the CG-regions of DNA; heterochromatic blocks are CG-rich, and thus offer a plethora of sites for CMA to bind to, intensifying the signal given off by chromomycin A3 (Mendes *et al* 2011). Conversely, these heterochromatic blocks are lacking in AT-rich regions, which DAPI more readily binds to (Mendes *et al* 2011). Chromomycin A3, a GC base-specific fluorochrome, counterstained with the AT-specific 4'-6-diamidino-2-phenyl-indole fluorochrome, revealed that most species and many Citrus cultivars are karyotypically distinct (Brasileiro-Vidal *et al* 2006). Recently there has been some success in identifying citrus chromosomes by combining fluorescent staining and in situ

hybridization with repetitive sequence probes and BAC clones (da Costa Silva et al., 2011). Badaeva et al were able to identify translocations in wheat chromosomes utilizing a staining technique known as C-banding (Badaeva *et al* 2007). Badaeva et al review how translocations give rise to unusual meiotic formations known as quadrivalents, which can be visualized using a fluorescent staining technique on chromosomes. A ring of four chromosomes, referred to as a quadrivalent or tetravalent (Rickards 1983), does not display normal segregation in meiosis and will often yield recombined gametes, which are partially aneuploid and nonviable (Badaeva *et al* 2007). It is therefore possible that seedlessness, a sign of problems in fertility, could be associated with chromosomal aberrations caused by radiation.

Retrotransposon Activation in Response to Radiation

The most common type of transposable element is the retrotransposon, which also occupies a large amount of plant genomes (Gupta *et al* 2004). Retrotransposons are a genre of transposable elements which self-propagate by reverse transcription of RNA intermediates. They are divided into two large groups: long terminal repeat (LTR) and the non-LTR retrotransposons; the LTR retrotransposons are further sub-classified into the Ty1-*copia* and the Ty3-*gypsy* groups, which differ from one another in the position order of their genes as well as in their degree of sequence similarity (Kumar *et al* 1999). The LTR regions at both ends of the retrotransposon contain the signals for transcription

initiation and termination. Some classes of retrotransposons are known to insert preferentially near or within genes (Gupta *et al* 2004) however, preferential integration sites for plant retrotransposons vary; in some cases sites are preferable in actively transcribed chromosomes and in others, sites are preferable in repeated heterochromatic regions (Kumar *et al* 1999). Retrotransposons can more readily generate phenotypic mutations by inserting within or near genes. Insertions in or next to coding regions generate mutations that can lead to changes in gene expression (Grandbastien *et al* 2005). Transposons can also inactivate genes by integration or by triggering methylation in the area where they are located (Biswas *et al* 2010). Retrotransposon-induced mutations are relatively stable because they are brought about through the introduction of a replicated DNA sequence; the sequence at the insertion site is retained. Stress factors have been shown to influence many plant retrotransposons (Grandbastien *et al* 2005); *Tto1* and *Tto2* of tobacco and *Tos17* of rice, isolated as transcriptionally active retrotransposons, have been demonstrated to be mobile in cultured cells (Hirochika *et al* 1993, Hirochika *et al* 1996). Retrotransposon transcription may be activated by stress factors such as wounding, biotic elicitors, pathogen attacks, and radiation exposure. These findings suggest that exposure to gamma radiation could activate various plant retrotransposons, which are inactive under normal growth conditions. Moreover, the activation of previously inactive retrotransposons could generate mutations within the DNA of the plant due to retrotransposon replication and insertion. Because many retrotransposons are widely distributed within the euchromatin domains of chromosomes, it should be possible to generate markers linked to agronomically important genes (Kumar *et al* 1999).

Although there are various methods one can use when studying retrotransposons within a genome, one popular method uses conserved regions within the retrotransposon as a template for designing the primers to amplify these regions in quantitative polymerase chain reaction (qPCR) (Gupta *et al* 2004). Quantitative PCR (qPCR) is the most accurate, precise, and efficient method currently available for DNA quantitation. It is an amplification method that detects and measures the accumulation of DNA that binds fluorescent dyes as the reaction progresses. The utilization of qPCR as a way of estimating the copy number of a genomic sequence of interest within a genome of interest has been proven successful in some studies (Ferreira *et al* 2006, Pickard *et al* 2003, Price *et al* 2004). Southern blot quantification and BLAST sequence alignment against reference genomes provide additional supporting evidence when estimating copy number. Activation of a retrotransposon in response to an abiotic stress factor, such as ionizing radiation, could lead to subsequent inactivation of a gene due to preferential insertion of retrotransposon copies close to or within genes. It is entirely possible that a gene important to the function of plant fertility could be affected by an interfering retrotransposon insertion, thus resulting in a plant's inability to produce seeds.

Epigenetic Regulation in Response to Radiation

DNA is packaged into chromatin that exists in two forms: euchromatin and heterochromatin. Heterochromatin, which is the more compact form of chromatin, has a

silent state associated with it that correlates with the presence of distinctive, repressive epigenetic modifications. In plants and mammals, heterochromatic DNA is densely methylated at cytosine residues (^{m}C). In plants, methylation is found at symmetric CpXpG and asymmetric CpXpX sites, as well as the canonical CpG sites found in most mammalian genomes (Rabinowicz *et al* 2003). Very high m^5C content (up to about 9 mol%) in total DNA is a specific feature of plants (Vanyushin *et al* 2006). Plant DNA is more heavily methylated than that of most other organisms (Vanyushi *et al* 2006). A variety of environmental stimuli affect DNA transcriptional responses; this has been documented for genes that are associated with euchromatin as well as for gene networks (Tittel-Elmer *et al* 2010). These responses have been associated with alterations in epigenetic regulatory mechanisms, such as changes in the distribution of DNA methylation. Work on animals, plants, and fungi has led to the conclusion that gene silencing is a key biological outcome of DNA methylation (Bird *et al* 1999). Cytosine methylation has been associated with multiple functions of molecular significance. Primarily, it is heavily involved in the alteration of gene expression (Cai *et al* 1995). This includes inhibition and activation of gene transcription through methylation and demethylation within promoter regions. Methylation and demethylation are believed to be key processes used in regulating cell- and organ-specific gene expression. Modifications in methylation patterns as well as other epigenetic marks at target genes seem to play vital roles in stress responses (Tittel-Elmer *et al* 2010). For instance, it is known that exposure to radiation can cause DNA methylation changes, and that these changes are dose-, sex- and tissue-specific (Tamminga 2008). Rapid progress has been

made in the development of highly sensitive and reliable methods for the identification and quantification of methylated cytosines in DNA (Razin *et al* 1980). One such application utilizes bacterial restriction endonucleases which cleave DNA at target sites, such as CpG islands, which are characteristically methylated. One can take advantage of these enzymes when probing for methylation because certain isoschizomers differ in ability to cut methylated DNA (Razin *et al* 1980). The Luminometric Methylation Assay (LUMA) exploits these enzymes by digesting DNA with enzymes that target the recognition sequence CCGG, which may contain a CpG methylation site. A pyrosequencing reaction is then used to fill in the overhangs which will be less frequent in DNA digested with the methylation-sensitive enzyme. Signal from this reaction measures the proportion of methylated targets in the DNA. Another application uses a methylation sensitive alteration of the amplified fragment length polymorphism (AFLP) technique to identify a subset of methylations in the genomes of organisms (Xiong *et al* 1999). This is a strong method for fingerprinting genomes that joins restriction fragment length polymorphism (RFLP) with the polymerase chain reaction (PCR). In methylation sensitive AFLP, isoschizomers HpaII and MspI tend to be used in parallel reactions. Each restriction enzyme recognizes the sequence CCGG but differs in its sensitivity to DNA methylation at the inner cytosine (Jeddeloh *et al* 1996). When the resulting amplified fragments are run side by side on a gel, a difference in band pattern indicates a methylation polymorphism. Changes in methylation patterns can cause significant changes in gene expression. It is entirely possible that a change in gene expression could

affect vital functions of a plant such as its fertility, thus resulting in seedlessness in the plant.

Dissertation Overview

The objective of this dissertation project was to evaluate and compare the molecular genetic differences between the seedless mandarin variety Tango with its seedy predecessor variety, W. Murcott. To determine the underlying molecular differences between W. Murcott and Tango citrus plants, molecular regions prone to alteration due to radiation exposure were targeted for analysis. The regions selected were whole chromosomal spreads from meiotic cells, LTR (long terminal repeat) retrotransposon regions, and methylated/unmethylated CpG regions within the genome. Due to the capability of each of these regions to undergo genetic alteration when put under stressful abiotic conditions (i.e. gamma radiation) it was hypothesized that each of these regions could prove genetically different between W. Murcott and Tango. To analyze these regions, the following experimental techniques were performed: for chromosomal spread analysis, images of meiotic metaphase chromosome spreads stained with chromomycin A3 and counterstained with DAPI from W. Murcott and Tango anther cells were captured via a Leica SP2 Confocal microscope. These images were compared and contrasted between the two citrus varieties to identify significant differences in chromosome alignment, as well as look for chromosomal alignment abnormalities that could be caused by aberrations such as translocations. For LTR retrotransposon quantitation, CIRE1, a fully sequenced, potentially active LTR retrotransposon in the *Citrus sinensis* genome,

was targeted for quantitative analysis in both the W. Murcott and Tango genomes. Primer sets were designed from the CIRE1 sequence found within the NCBI database archives, and targeted three specific regions of CIRE1: a long terminal repeat region, an integrase gene, and a reverse transcriptase gene. The most conserved region of the three, the LTR region, was quantified using qPCR real-time amplification while all three regions were quantified using Southern blot and BLAST sequence alignment. For CpG methylation analysis, the proportion of methylated CpG sequences in CCGG motifs was estimated using LUMA analysis of W. Murcott and Tango DNA extractions from young leaf and gynoecium tissue. Methylation-sensitive Amplified Fragment Length Polymorphism (AFLP) analysis was also performed by Lisa Mu to determine whether or not differences in methylation patterns between W. Murcott and Tango could be visualized. Restriction enzymes *EcoRI* and *MseI* were used in combination to digest W. Murcott and Tango genomic DNA, with *EcoRI* being a frequent cutter enzyme and *MseI* being a methylation sensitive enzyme. Adapters of known sequence were ligated to the severed DNA ends and polymerase chain reaction (PCR) was used to amplify these regions. The amplified regions were size fractionated by polyacrylamide gel electrophoresis and patterns were compared for differences in banding size (indicating differences in methylated sites within the genomic DNA).

Literature Cited

Anonymous. 1960. Radiation and crop improvement. IAEA Bulletin 2 (4): 8-9

Badaeva, E.D., Dedkova, O.S., Gay, G., Pukhalskyi, V.A., Zelenin, A.V., Bernard, S., and Bernard, M., 2007. Chromosomal rearrangement in wheat: their types and distribution. Genome 50: 907-926.

Barrett, H.C., and Rhodes, A.M., 1976. A numerical taxonomic study of affinity relationships in cultivated Citrus and its close relatives. Syst. Bot. 1: 105–136.

Bird, A.P., and Wolffe, A.P., 1999. Methylation-induced repression--belts, braces, and chromatin. Cell 99: 451–454.

Biswas, M.K, Baig, M.N.R., Cheng, Y.-J., and Deng, X.-X., 2010. Retro-transposon based genetic similarity within the genus *Citrus* and its relatives. Genet. Resour. Crop Evol. 57: 963–972.

Brasileiro-Vidal A.C., dos Santos-Serejo J.A., dos S. Soares Filho W. and Guerra M., 2006. A simple chromosomal marker can reliably distinguishes *Poncirus* from *Citrus* species. Genetica 129: 273–279.

Cai, Q., Guy, C.L., and Moore, G.A., 1995. Detection of cytosine methylation and mapping of a gene influencing cytosine methylation in the genome of Citrus. Genome 39: 235-242.

Carvalho R., Soares Filho W.S., Brasileiro-Vidala A.C. and Guerraa M., 2005. The relationships among lemons, limes and citron: a chromosomal comparison. Cytogenet. Genome. Res. 109: 276–282.

Chopra, V.L., 2005. Mutagenesis: Investigating the process and processing the outcome for crop improvement. Current Science 89: 353-359.

da Costa Silva, S., Marques, A., dos Santos Soares Filho, W., Mirkov, T.E., Pedrosa-Herand, A., and Guerra, M., 2011. The cytogenetic map of the *Poncirus trifoliata* (L.) Raf.—a nomenclature system for chromosomes of all citric species. Tropical Plant Biol. 4:99–105.

de Vozzi M.S., Santos S.A., Pereira C.S., Cuzzi J.F., Laureano L.A., Franco Jr. J.G. and Martelli L., 2009. Meiotic segregation and interchromosomal effect in the sperm of a double translocation carrier: a case report. Molecular Cytogenetics 2: 24.

Ferreira, I.D., do Rosário, V.E., and Cravo, P.V.L., 2006. Real-time quantitative PCR with SYBR Green I detection for estimating copy numbers of nine drug resistance candidate genes in *Plasmodium falciparum*. Malar. J. 5: 1.

Findley, S.D., Cannon, S., Varala, K., Du, J., Ma, J., Hudson, M.E., Birchler, J.A., and Stacey, G., 2010. A fluorescence in situ hybridization system for karyotyping soybean. Genetics 185: 727–744.

Gmitter, F. and Hu, X., 1990. The possible role of Yunnan, China, in the origin of contemporary *Citrus* species (*Rutaceae*). Econ. Botany 44: 267–277

Grandbastien, M.-A., Audeon, C., Bonnivard, E., Casacuberta, J.M., Chalhoub, B., Costa, A.-P.P., Le, Q.H., Melayah, D., Petit, M., Poncet, C., Tam, S.M., Van Sluys, M.-A., and Mhiria, C., 2005. Stress activation and genomic impact of *Tnt1* retrotransposons in *Solanaceae*. Cytogenet. Genome Res. 110:229–241.

Gupta, P.K., and Rustgi, S., 2004. Molecular markers from the transcribed/expressed region of the genome in higher plants. Funct. Integr. Genomics 4: 139–162.

Hearn, C.J., 1984. Development of seedless orange and grapefruit cultivars through seed irradiation. J. Amer. Soc. Hort. Sci. 109: 270-273.

Hirochika, H, 1993. Activation of tobacco retrotransposons during tissue culture. EMBO J. 12: 2521–2528.

Hirochika, H, Sugimoto, K, Otsuki, Y, Tsugawa, H, and Kanda, M, 1996. Retrotransposons of rice involved in mutations induced by tissue culture. Proc. Natl. Acad. Sci. USA 93: 7783–7788.

Huang, J.C., Xiao, Y., Zhao, C.X., Chen, Z.S., and Wang, X.Y., 2003. Induction of superior seedless mutation of *C. grandis* O*sbeck cv*.*Sshatianyou* by irradiation. Acta Agric. Nucleatae Sin. 17: 171–174.

Jeddeloh, J. A., and E. J. Richards, 1996. mCCG methylation in angiosperms. Plant J. 9: 579-586.

Kahn, T.A. and Chao, C.T., 2004. Mysteries of mandarins. Sex, seedlessness and new varieties. UC Cooperative Extension C 662: 1-4. *cesandiego.ucdavis.edu/files/54349.pdf*

Kamm, A., Galasso, I., Schmidt, T., and Heslop-Harrison, J.S., 1995. Analysis of a repetitive DNA family from *Arabidopsis arenosa* and relationships between *Arabidopsis* species. Plant Mol. Biol. 27: 853–862.

Karl Sax 1938. Chromosome aberrations induced by X-rays. Genetics 3: 494.

Kato, A., Lamb, J.C., and Birchler, J.A., 2004.Chromosome painting using repetitive DNA sequences as probes for somatic chromosome identification in maize. Proc. Nat. Acad. Sci. U.S.A. 101: 13554–13559.

Kumar, A., and Bennetzen, J.L., 1999. Plant retrotransposons. Annu. Rev. Genet. 33: 479-532.

Lawton, H.W. and Weathers, L.G., 1989. The origins of citrus research in California. In: Reuther W., Calavan E.C., and Carman G.E. (eds.) The Citrus Industry, Volume 5. The Regents of the University of California Division of Agriculture and Natural Resources, 281-335.

Matsui A, Ishida J, Morosawa T, Mochizuki Y, Kaminuma E, et al, 2008 Arabidopsis transcriptome analysis under drought, cold, high-salinity and ABA treatment conditions using a tiling array. Plant Cell Physiol. 49: 1135–1149.

Mendes, S., Moraes, A.P., Mirkov, T.E, and Pedrosa-Harand, A., 2011. Chromosome homeologies and high variation in heterochromatin distribution between *Citrus* L. and *Poncirus* Raf. as evidenced by comparative cytogenetic mapping. Chromosome Res. 19: 521-530.

Natarajan, A.T., 2005. Chromosome aberrations: Plants to human and Feulgen to FISH. Current Science 89: 335-340.

Ollitrault, P., Froelicher, Y., Dambier, D., Luro, F., and Yamamoto, M., 2007. Seedlessness and ploidy manipulation. In: Citrus Genetics, Breeding, and Biotechnology (ed. I.A. Khan), CAB International, Wallingford, U.K., 197-218.

Pang, X-M., Hu, C-G., and Deng, X-X, 2007. Phylogenetic relationships within *Citrus* and its related genera as inferred from AFLP markers. Genetic Resources and Crop Evolution 54: 429–436.

Pickard, A.L., Wongsrichanalai, C., Purfield, A., Kamwendo, D., Emery, K., Zalewski, C., Kawamoto, F., Miller, R.S., and Meshnick, S.R., 2003. Resistance to antimalarials in southeast Asia and genetic polymorphisms in *pfmdr1*. Antimicrob. Agents Chemother. 47: 2418–2423.

Price, R.N., Uhlemann, A.C., Brockman, A., McGready, R., Ashley, E., Phaipun, L., Patel, R., Laing, K., Looareesuwan, S., White, N.J., Nosten, F., and Krishna, S., 2004. Mefloquine resistance in *Plasmodium falciparum* and increased *pfmdr1* gene copy number. Lancet. **364**: 438–447.

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

Raza, H., Khan, M.M., and Khan, A.A., 2003. Review: seedlessness in Citrus. Int. J. Agri. Biol. 5: 388-391.

Razin, A., and Riggs, A.D.,1980. DNA methylation and gene function. Science 210: 604-610.

Rickards, G.K., 1983. Orientation behavior of chromosome multiples of interchange (reciprocal translocation) heterozygotes. Ann. Rev. Genet. 17: 443-498.

Roose M.L., Schwarzacher T., and Heslop-Harrison J.S., 1998. The chromosomes of *Citrus* and *Poncirus* species and hybrids: identification of characteristic chromosomes and physical mapping of rDNA loci using in situ hybridization and fluorochrome banding. J. Heredity 89: 83-86.

Roose, M.L. and Williams, T.E., 2007a. Mutation breeding. In: Citrus Genetics, Breeding, and Biotechnology (ed. I.A. Khan), CAB International, Wallingford, U.K., 345-352.

Roose ML and Williams TE. 2007b. Mandarin tree named 'Tango'. US Plant Patent PP17,863.

Savage, J.R.K. and Simpson, P.J., 1993. FISH "painting" patterns resulting from complex exchanges. Mutation Research 312: 51-60.

Scora, R.W., 1975. On the history and origin of Citrus. Bull. Torrey Bot. Club 102: 369–375.

Sonah, H., Deshmukh, R.K., Singh, V.P., Gupta, D.K., Singh, N.K., and Sharm, T.R., 2011. Genomic resources in horticultural crops: Status, utility and challenges. Biotechnology Advances 29: 199–209.

Sutarto, I., Agisimanto, D., and Supriyanto, A., 2009. Development of promising seedless citrus mutants through gamma irradiation. Induced Plant Mutations in the Genomics Era. Food and Agriculture Organization of the United Nations, Rome 306-308.

Tamminga, J., 2008. Radiation-Induced Epigenome Deregulation in the Male Germline. Department of Biological Sciences University of Lethbridge 1-102.

Terol, J., Naranjo, M.A., Ollitrault, P., and Talon, M., 2008. Development of genomic resources for *Citrus clementina*: Characterization of three deep-coverage BAC libraries and analysis of 46,000 BAC end sequences. BMC Genomics 9: 423.

Tittel-Elmer, M., Bucher, E., Broger, L., Mathieu, O., Paszkowski, J., Vaillant, I., 2010. Stress-induced activation of heterochromatic transcription. PLoS Genetics 6(10): e1001175. doi:10.1371/journal.pgen.1001175 6: 1-11.

Vanyushin, B.F., 2006. DNA methylation in plants. CTMI 301: 67–122.

Van Harten, A.M., 1998. Mutation Breeding: Theory and Practical Applications. Cambridge University Press, Cambridge.

Vardi, A, Elhanati, A, Frydman-Shani A., and Neumann H. (1996) Strategies and considerations in mandarin improvement programmes. Proc. Int. Soc. Citriculture 1: 109-112.

Viccini L.F. and de Carvalho C.R., 2002. Meiotic chromosomal variation resulting from irradiation of pollen in maize. J. Appl. Genet. 43(4): 463-469.

Webber, H.J., Reuther, W., and Lawton, H.W., 1967. History and development of the citrus industry. In: The Citrus Industry, Vol. I (ed. W. Reuther, H. J. Webber and L D. Batchelor, University of California Division of Agricultural Sciences, 1-39.

Williams, T.E., and Roose, M.L., 2010. 'Tango—a new, very low-seeded, late-season irradiated selection of 'W. Murcott' mandarin from the University of California Riverside. Proc. Int. Society of Citriculture 1: 202.

Wright G.C., 2007. Pollination of W. Murcott Afourer mandarins. Citrus Research Report (P-153).

Xiong, L. Z., C. G. Xu, M. A. Saghai Maroof, and Q. Zhang. 1999. Patterns of cytosine methylation in an elite rice hybrid and its parental lines, detected by a methylation sensitive amplification polymorphism technique. Mol. Gen. Genet. 261: 439-446.

Ye W., Qin Y., Ye Z., Teixeira da Silva J.A., Zhang L., Wu X., Lin S. and Hu G., 2009. Seedless mechanism of a new mandarin cultivar 'Wuzishatangju' (*Citrus reticulata* Blanco). Plant Science 177: 19–27.

Zeller G, Henz SR, Widmer CK, Sachsenberg T, Ratsch G, et al, 2009. Stress induced changes in the Arabidopsis thaliana transcriptome analyzed using whole-genome tiling arrays. Plant J. 58: 1068–1082.

.

Chapter 2

Comparison of Meiotic Chromosome Pairing in W. Murcott Mandarin and its Irradiated Offshoot, Tango

Abstract

Seedless citrus fruit can be commercially valuable due to its high demand by consumers. For the convenience of not having to dispose of seeds when consuming citrus fruit, consumers are willing to pay a premium price. Citrus breeders therefore have an incentive to find new ways of developing seedless fruit cultivars. In mandarin cultivars, cultivars producing seedless fruit can be developed by mutation breeding that induces chromosomal rearrangements. These rearrangements cause abnormal meiosis resulting in gametes with duplications and deficiencies and embryo abortion. The purpose of this study was to analyze and compare meiotic chromosome spreads taken from the anther cells of W. Murcott, a seedy mandarin, and from Tango, a seedless mandarin derived from W. Murcott through an irradiation process. Meiotic chromosome spreads were stained with chromomycin A_3 (CMA3) and counterstained with 4,6-diamidino-2-phenylindole (DAPI) in order to try to visualize differentiation between the chromosomes by targeting their heterochromatic block regions. All photographed and documented meiotic chromosome spreads taken of W. Murcott showed normal chromosomal alignment in metaphase and normal chromosomal segregation in anaphase. All photographed and documented meiotic chromosome spreads taken of Tango showed

abnormal chromosomal alignment in metaphase and abnormal chromosomal segregation in anaphase. Moreover, some meiotic chromosome spreads in Tango contained quadrivalents, and others lagging chromosomes. Both quadrivalents and laggard chromosomes in meiosis may lead to pollen sterility. The results indicate that male sterility in Tango is likely based, at least in part, on chromosome rearrangements. In contrast, no significant deviances from normal chromosomal alignment and separation were found in W. Murcott. This result implies that in W. Murcott microsporocyte development is normal, an observation consistent with its high pollen viability.

Introduction

Citrus cultivars are acknowledged as one of the world's most important food crops (Hernandez-Montoya *et al* 2009). They are cultivated in many countries having a tropical or subtropical climate; the foremost producers being Brazil, China, Japan, Mexico, Pakistan, USA and countries of the Mediterranean region (Hernandez-Montoya *et al* 2009). *Citrus* trees are an economically important fruit crop, with an annual worldwide production surpassing 11 million tons in 2009 (FAOSTAT 2011). The economically significant citrus fruits stem from natural hybridization and backcrossing amongst *Citrus maxima* [Burm.] Merrill (pummelo), *C. reticulata* Blanco (mandarin), and *C. medica* L. (citron) (Mabberley 1997). A good amount of the hybrids procreate by a method of apomixis—known as nucellar embryony—and have been appointed species names by a number of authorities (Roose *et al* 1998). Commercial citrus cultivars are genetically linked with several species which are propagated through asexual means by grafting onto rootstock cultivars (Brasileiro-Vidal *et al* 2006).

Production of seedless fruit is a valuable trait in fresh fruit cultivars in that it carries a strong correlation with its economic value (Chao *et al* 2005). Seedless fruit have a premium price because consumers demand the convenience of not having to dispose of seeds. Therefore, development of seedless fruit cultivars has become a primary objective for citrus fruit breeders across the globe (Ye *et al* 2009). One particular variety of citrus, a seedless mandarin-producing plant known as Tango, originates from a predecessor variety, a seedy mandarin-producing plant known as W. Murcott. The Tango variety was

developed by propagation of trees from gamma-irradiated W. Murcott budwood and selection of a tree producing very low-seeded fruit. In the case of Tango, fruit produced by a mature tree retained the same characteristics of size, color, easiness to peel, and sweetness as that produced by W. Murcott; Tango fruit differed from W. Murcott fruit only in its seedlessness (Roose and Williams 2007).

Cross-pollination can greatly amplify the number of seed per fruit in many citrus crops (Chao *et al* 2005). In situations where there are compatible pollen sources and pollinators, the utmost seed number per fruit obtained from field sampling is comparable to that of hand cross-pollination (Chao *et al* 2005). An appropriate distance, which can range up to a few kilometers, is required between two compatible citrus cultivars to circumvent cross-pollination by bees and to ensure seedless fruit production (Chao *et al* 2005). W. Murcott, for example, has increased seediness when cross-pollinated with other *Citrus* cultivars. Tango, however, maintains a relatively seedless state (less than 0.2 seeds per fruit, on average) even when cross-pollinated with other citrus varieties (Roose and Williams 2007). This capability makes Tango exceptionally commercially desirable as it can be planted within the vicinity of other citrus varieties without risk of substantial seediness of its fruit. An additional advantage of Tango is that it produces little viable pollen and therefore will not cause seediness in cultivars located nearby.

Generation of seedless fruit in mandarin cultivars can be brought about in several different ways (Ye *et al* 2009). Seedlessness can be brought about by deterioration of anther, pollen grains, and microsporocyte or chromosomal rearrangements that trigger

aberrant meiosis in microsporocytes, as well as extreme climatic changes (Ye et al 2009). Seedless fruit production is a combination of parthenocarpy (i.e., the formation of fruit devoid of fertilization or embryo abortion) with female sterility and/or male sterility, embryo sac abortion, self-incompatibility (SI), polyploidy, abnormal climate, and treatment with plant growth regulators such as gibberellins (Ye *et al*, 2009). Bud variation is one type of somatic mutation that can produce seedless fruit. Numerous seedless fruit cultivars, in particular seedless mandarin cultivars, have been chosen from bud variation of seedy cultivars (Shamel 1943, Hensz 1982, Ye *et al*, 2009). The source of bud variation stems from alteration in the genetic structure, including changes in the number of chromosomes per cell, irregularity in chromosome structure and gene mutations (Ye *et al* 2009). Furthermore, bud variation from gene mutations is generally monogenically controlled (Ye *et al* 2009).

Male sterility in *Citrus* is normally defined as pollen abortion and it strongly correlates with the seedless phenotype of a fruit. Findings have indicated that chromosomal variation and abnormal segregation during meiosis of the microsporocyte could result in anther sterility in citrus (McClintock 1984, Marcucci *et al* 1984, Viccini and de Carvalho 2002, Ye *et al* 2009). Single breaks in the chromosome have been found to generate an acentric fragment, while partial breakage may generate an achromatic lesion (Sax 1938) meaning that a full break is required in order to release the distal end of the affected chromosome, while a partial break will produce a chromosomal lesion but will not release the distal end of the affected chromosome. Double breaks have been found to cause a reciprocal exchange of chromotid fragments, that is, broken fragments from

individual chromatids are switched between one another and are subsequently fused onto the reciprocating chromatid (Sax 1938). This chromatid interchange is nearly always equal, although instances where there is an unequal chromatid exchange have also occurred, albeit less frequently.

Evans and Bigger noted that root-tip cells in *Vicia faba* were found to have such interchange between chromatids when exposed to gamma radiation (Evans and Bigger 1961). Sanamyan and Rakhmatullina also reported such interchanges between chromatids in cotton plants induced through gamma-irradiation technique (Sanamyan and Rakhmatullina 2003); they noted that the rate of single exchanges was shown to be higher in gametic cells from gamma-irradiated pollen. Aneuploidies may result from meiotic-segregation errors, non-disjunction owing to recombination defects, and changes in kinetochore and microtubule structure (de Vozzi *et al* 2009). DNA breakage must occur in order for rearrangements to take place in other chromosomes, creating structural abnormalities (de Vozzi *et al* 2009). During the process of meiotic recombination, chromosomes are particularly vulnerable to loss or gain of genetic material should a breakage in DNA occur (de Vozzi *et al* 2009). Chromosomal rearrangements have played a key role in the evolution of plants (Badaeva *et al* 2007). The introduction of new species oftentimes is facilitated through chromosomal translocations, inversions, as well as other structural chromosome changes (Badaeva *et al* 2007). Reciprocal translocation is the most common type of translocation found in natural populations and refers specifically to the exchange of segments between chromosomes (Rickards *et al* 1983). Understanding translocation types, including the rate at which they occur and their

environmental distribution, is crucial to recognizing the evolutionary significance of chromosomal rearrangements and how they affect speciation and intraspecific divergence (Badaeva *et al* 2007). The detection of chromosomes that tend to engage in such translocations, as well as the location within the chromosome that is prone to breakage, will offer useful information on the processes involved in chromosomal rearrangements and the circumstances, which stimulate their emergence and fixation (Badaeva *et al* 2007).

Aberrant meiosis, which is a problematic outcome of reciprocal translocations, might produce gametes lacking portions of chromosomes (Curtis *et al* 2008). The exclusion of genes that are located on the removed chromosome regions could adversely affect gametophyte development (Curtis *et al* 2008). In meiotic cell division, translocations give rise to quadrivalents; segregation does not happen normally in such chromosomal arrangements, resulting in both parental and recombined gametes, with the latter being partially aneuploid and, consequently, nonviable (Badaeva *et al* 2007). An interchromosomal effect (ICE) is defined as "the influence of translocated chromosomes on the synapsis and disjunction of other chromosomes" (de Vozzi *et al* 2009). ICE has been identified in a number of chromosomal rearrangements (de Vozzi *et al* 2009). Chromosomal analysis of male gametic cells from organisms displaying low or no fertility have shown high variability in chromosomal segregation behaviours during meiosis (de Vozzi *et al* 2009).

Findings by Curtis *et al* suggest that a common post-meiotic checkpoint could discern the introduction of chromosome duplications or deletions, thus safeguarding against the

inclusion of atypical gametes within pre-fertilization pools. (Curtis *et al* 2008). A study by Henry et al (2007) proposes that there exists a likely genetic modifier of aneuploid survival, known as the Sensitive to Dosage Imbalance (SDI) locus. It is unknown at what stage of haploid or diploid development SDI engages in aneuploid survival (Curtis *et al* 2008).

There are different cytogenetic approaches for the detection of translocations. C-banding is both an inexpensive and effectual approach for wide-scale evaluation of some species of plants. This makes it possible for researchers to identify the genetic structure of varieties relatively quickly, as well as pinpoint major chromosomal aberrations and determine where breakpoints on chromosomes are located (Badaeva *et al* 2007). However, C-banding results such as this require that large segments of the chromosome be involved in such chromosomal rearrangements (Badaeva *et al* 2007) or alter the heterochromatin pattern of distribution on chromosomes.

Rearrangements involving small chromosomal segments and chromosomal segment exchanges between chromosomes that lack prominent C-bands may not be visible (Badaeva *et al* 2007). Previous studies and data suggest that many small chromosomal rearrangements that are not detected in mitotic observation can be seen in meiotic analyses using classical or Giemsa staining or by fluorescent in situ hybridization (FISH) and genomic in situ hybridization (GISH) (Badaeva *et al* 2007). Homologous pairing of chromosomes may be the most dependable means for distinguishing any alteration in chromosome structure, and could be aided through the use of in situ hybridization

(Badaeva *et al* 2007). The in situ-hybridization (ISH) technique, which was developed over 30 years ago, allows for sequences of DNA to be directly pinpointed on chromosomes in cytological preparations (Heslop-Harrison *et al* 1998). The introduction of user-friendly fluorescent techniques has attracted more researchers into taking advantage of this technique during in the past 15 years (Heslop-Harrison *et al* 1998). FISH allows for the exact site of hybridization to be detected visually and, additionally, it allows for a number of probes to be used at one time so they may be simultaneously visualized with different fluorochrome (Heslop-Harrison *et al* 1998).

Nearly all *Citrus* species are diploid, with $2n = 2x = 18$ chromosomes, and most are outcrossing and highly heterozygous. No inbred lines have been developed (Roose *et al* 1998). The genome size is small at about 382 Mbp (Roose *et al* 1998). Citrus cytogenetics has been limited by the small and relatively similar chromosomes (Roose *et al* 1998). Despite having a stable chromosome number ($2n = 18$) as well as being similar in chromosome size and morphology, the *Citrus* species and its closely related genera vary in the number and location of heterochromatic blocks within chromosomes (Carvalho *et al* 2005). These block variations can be easily observed when stained with fluorochromes chromomycin A3 (CMA3) and 4,6-diamidino-2-phenylindole (DAPI) using a double-staining technique (Carvalho *et al* 2005).

Another technique, known as multicolor fluorescent in situ hybridization, provides a useful approach to the problem of chromosome analysis and identification (Roose *et al* 1998). Such technology has been applied in important cereal species and many other

species, including those with small chromosomes such as *Arabidopsis thaliana* (Roose *et al* 1998). There have been several types of chromosomes that have been identified in *Citrus* as well as related genera by chromomycinA3 (CMA3) banding patterns in conjunction with FISH (Brasileiro-Vidal *et al* 2006). CMA3, which is a fluorochrome specific to base-pairs GC, counterstained with the AT-specific 4'-6-diamidino-2-phenyl-indole (DAPI) fluorochrome, was found to karyotypically differentiate the majority of *Citrus* cultivars as well as a number of other plant species (Brasileiro-Vidal *et al* 2006).

It is the intent of this study to examine, compare and contrast chromosomal spreads taken from *W. Murcott* and *Tango* mandarin trees to determine whether or not chromosomal structural abnormalities that result in abnormal meiotic pairing might be responsible for the seedless characteristic of *Tango* fruit; it is hypothesized that abnormal chromosomal formations during the metaphase and anaphase stages of meiosis would result in the formation of gametes with chromosomal abnormalities such as aneuploidy, thus rendering the gametes inviable and the affected tree sterile.

Materials and Methods

Tissue Collection and Preservation

Immature flower buds, sized 0.5-2 mm in diameter, were collected from UCR research citrus fields and greenhouses. Samples were taken from three separate Tango mandarin trees and two separate W. Murcott mandarin trees. Flower buds were kept in plastic bags on ice until ready for further treatment. To freeze cells in a metaphase/anaphase state, whole flower bud tissue was submerged in several volumes of 8-hydroxyquinoline (0.002 M) for 24 hours at 8°C. Tissue was then fixed in an ethanol-acetic acid fixative (3:1 ratio, respectively) overnight at room temperature, followed by storage in fixative at -20°C.

To digest tissue for squashing, an immature flower bud was extracted from fixative using sterilized laboratory forceps. A razor was used to slice open the bud, whereupon the anthers within the flower bud were removed with the forceps under a dissecting microscope. Anthers were placed in a microcentrifuge tube containing several microliters of a digestion solution [2% cellulose from *Aspergillus niger*-20% pectinase in 0.01M citrate buffer, (pH 4.0)] and were heated at 37°C for at least 3 hours to as long as overnight. This protocol for tissue collection and preservation was followed according to Carvalho et al (2005).

Chromosome Spread Preparation

Chromosome sprads were prepared as described previously for rye (Masoudi-Nejad *et al* 2002). Glass slides were sterilized in ethanol by inserting into a glass-slide holder tube

containing several ml of 100% ethanol. Slides were prepared for chromosome spread by wiping off all liquid using a kimwipe and setting aside. Anther plant tissue that had been softened by cellulase/pectinase enzyme mixture was extracted from the plastic microcentrifuge tube using sterilized laboratory forceps; a drop of 45% acetic acid was applied to the glass slide, approximately two anthers were then placed on the slide, the anthers in the drop of acetic acid were covered with a small glass coverslip and were "squashed" by tapping gently with a match until the tissue was disintegrated into a ring of loose cells. Once no fragments of intact tissue remained, the slide was heated briefly by clasping one end of the slide with a pair of sterilized laboratory forceps and waving the slide over a Bunsen burner flame several times for a few seconds, being careful not to burn the cells by allowing the flame to touch the glass slide. The slide was allowed to cool for several seconds before pressing down on the coverslip to help further spread out cells. The slide was then placed, coverslip side up, on a large piece of dry ice, and was allowed to freeze for around ten minutes (meanwhile, more slide "squashes" were made and also placed on the large piece of dry ice, so that an accumulation of several slides could be placed on the ice at a time).

After allowing the coverslip to freeze to its slide, a razor blade was inserted underneath one corner of the coverslip; the coverslip was subsequently removed from the slide by lifting the razor blade underneath the coverslip corner to overturn the coverslip. The slide was then immediately submerged in 45% acetic acid at room temperature for at least one minute, followed by submersion in 45% acetic acid at 48-53°C for three minutes. The slide was removed from warmed acetic acid and allowed to air dry. Slides prepared in

this manner were either used immediately or stored at room temperature until ready for use, no more than one week.

DAPI/CMA3 Staining

Procedures for DAPI/CMA3 staining followed the CMA3 stain protocol described by Nasr-Esfahani et al (2009) and the DAPI counterstain protocol described by ???? Cells were fixed onto glass slides by immersing in ice-cold 95% EtOH/5% glacial acetic acid for ten minutes. Slides were removed from fixative and were washed five times, at two minutes per wash, in 1X Phosphate-buffered saline (PBS) buffer (pH 7.4). The recipe used for making 1 X PBS buffer was as follows: 137 mM NaCl, 2.7 mM KCl, 100 mM Na_2HPO_4, and 2mM KH_2PO_4; pH to 7.4 using HCl. Cells were then permeabilized for staining by immersing the slides in 0.5% TritonX-100 in 1X PBS buffer for ten minutes. The slides were removed from this liquid after ten minutes and were rinsed three times, at five minutes per wash, in 1X PBS buffer. Slides were then removed from buffer and were ready for staining.

The following staining procedures were performed in the dark, due to light sensitivity of the dyes involved. Chromomycin A3 was prepared by adding 1 ml of 100% pure ethanol to a vial containing 5 mg of chromomycin A3 (AG Scientific, Inc.). Stock solution of 5mg/ml was diluted to working concentration of 0.5 mg/ml in McIlvaine's buffer (7 ml 0.1M citric acid, 32.9 ml 0.2M $Na_2HPO_4.7H_2O$, pH to 7.0, containing 10 mM $MgCl_2$). This working concentration was used directly on the slide specimens. About 100 µl was

placed on each slide, a plastic coverslip was used to distribute the chromomyacin A3 dye evenly along the glass slide. Slides were incubated in the dark for at least 1 hour.

4',6-diamidino-2-phenylindole (DAPI) (from Sigma-Aldrich) was prepared by diluting in double-distilled H_20 to a stock concentration of 100 µg/ml, followed by further dilution in 1X PBS buffer to a working concentration of 0.5 µg/ml. After incubation with chromomycin A3, slides were rinsed once in 1X PBS buffer for five minutes in coplin jars to remove unbound excess dye. The slides were then counterstained with DAPI by applying 300 µl of 0.5 µg/ml DAPI to each slide; a new plastic coverslip was applied to each slide to distribute the dye evenly. Slides were left to incubate in the dark for 30 minutes. After incubation with DAPI, slides were rinsed once in 1X PBS buffer for five minutes in coplin jars to remove unbound excess dye, followed by a change of fresh 1X PBS buffer and a second, final five minute rinse in 1XPBS buffer. Excess liquid around the sample on each slide was blotted with a kimwipe before applying 20 µl of antifade solution (McIlvaine's buffer and glycerol in a 1:1 ratio) and covering with a glass coverslip. Slides were carefully covered in a container to prevent light exposure, and were stored in a -20°C freezer until ready for viewing.

Leica SP2 Microscopy Used for Viewing Slides and Capturing Chromosomal Images

Chromosome spreads were analyzed and images were captured by a Leica SP2 confocal microscope, at the Microscopy Imaging Core at the University of California in Riverside (UCR). Images captured by the Leica SP2 confocal microscope were obtained with

assistance provided by Dr. David Carter. For DAPI excitation, a light wavelength filter of 358 nm (excitation maximum) to 461 nm (emission maximum) was used. For CMA3 excitation, a light wavelength filter of 445 nm (excitation maximum) to 575 nm (emission maximum) was used. X/Y coordinates were documented for each chromosomal spread analyzed as well as for each image taken of a particular spread. Colors were arbitrarily assigned to differentiate DAPI from CMA3 in the images taken; generally speaking, CMA3 was given a brighter color than DAPI as its light emission was often weaker and thus harder to detect.

Results

Metaphase and Anaphase Chromosome Spread Analysis

It proved challenging to isolate the ideal chromosomal spreads with the following characteristics: correct chromosomal number (2*N*=18), chromosomes in condensed state, and correct phase of meiosis (either metaphase or anaphase was acceptable). Of the many slides that were viewed, the average number of acceptable chromosomal spreads per slide was low (*Table 2.1*) and, in some cases, good spreads were evident but unfortunately were not captured by Leica SP2 confocal imaging due to DAPI/CMA3 bleaching caused by overexposure to light. However, a total of 69 acceptable images (*Table 2.2*) provided evidence of contrasting chromosomal alignment and segregation behavior in the metaphase and anaphase stages of meiosis in W. Murcott and Tango gametic cells, respectively. Chromosomal alignment and segregation behavior was described in detail and descriptive categories were assigned to each image based on the image of the chromosomes displayed.

Chromosome Spread Images

Several descriptive categories were created and each Leica SP2 confocal image was assigned to one of the following categories: Normal Alignment, Normal Segregation, Misalignment, Partial Misalignment, Quadrivalent, Aneuploidy, Lagging Chromosome, or Other. Chromosome spreads that seemed to display unique characteristics unlike the specified descriptive categories were automatically placed in the "Other" category. Chromosomal spreads of W. Murcott were predominantly in the metaphase stage of

meiosis (*Tables 2.2, 2.3*) and were categorized as being in Normal Alignment, as depicted in captured Leica SP2 confocal images (*Figure 2.1b, c*). Fewer spreads could be classified as being in the anaphase stage of meiosis (*Figure 2.1a, Tables 2.2, 2.3*) but these were classified into the Normal Segregation category. Rarely, a spread was seen that could not definitively be classified as either a metaphase or an anaphase spread (*Tables 2.2, 2.3*); more than likely, these spreads lay within an intermediate stage of meiosis such as prometaphase. These hard to specify spreads, such as that shown in Figure 2.1D, were placed in the Other category. Like W. Murcott, the majority of chromosomal spreads in Tango were found to be predominantly in metaphase (*Tables 2.2, 2.4*). However, unlike W. Murcott, chromosome spreads in metaphase were completely misaligned (*Figure 2.2a, b*) or at least partially misaligned (*Figure 2.2c, d*). Fewer spreads were found to be in the anaphase stage (*Tables 2.2, 2.4*) Chromosomes in the anaphase spreads of Tango did not display normal segregation (*Figure 2.2e, f*), implying that resulting gametes would be aneuploid. Lagging chromosomes were also evident in some chromosomal segregations captured (*Figure 2.2f*). The reason for chromosomal misalignment could sometimes be attributed to the presence of quadrivalents (*Figure 2.2a, g*). Some spreads were also classified as belonging to the "Other" category (*Tables 2.2, 2.3*) as depicted in captured Leica SP2 confocal images (*Figure 2.2g*); these spreads were generally scattered enough as to generate confusion towards placing them in either the metaphase or anaphase categories, however, their tendency to have high numbers of quadrivalents made them interesting enough to include in the chromosomal analysis data.

CMA3 Banding Patterns

Chromomyacin A3 banding rarely provided a sufficient amount of contrast for differentiation of the eight distinctive chromosomes found in *Citrus*. Chromomycin A3 is a weaker dye than is DAPI, and thus is more light sensitive and prone to bleaching despite application of copious amounts of antifade. Moreover, the stronger, more durable DAPI fluorescent signal often overpowers that of chromomycin A3, which also makes CMA3 banding patterns hard to detect. However, certain regions of the chromosome had more affinity to chromomycin A3 binding than others (*Figure 2.3*) as depicted by the contrasting brightness of CMA3 fluorescence against DAPI. Due to the small size of *Citrus* chromosomes (less than 5 µm) and the inability of the Leica SP2 confocal microscope to zoom into a close-up of chromosomes without also creating a blurred image due to graininess, specific chromosomes were unable to be identified with complete certainty based on CMA3 banding patterns.

Discussion

These results indicate that abnormal chromosomal alignment in the metaphase/anaphase stages of meiosis occur in Tango mandarin gametic cells, whereas such abnormal alignment is absent in W. Murcott mandarin gametic cells. In Tango, perfect linear alignment of homologous chromosomes during metaphase is shown to deviate from the norm, with some chromosomes arranged in circular formations, much like quadrivalents. According to Curtis et al (2008) spreads of chromosomes in meiosis I that display this cruciform-like formation are likely to have been subjected to a reciprocal translocation. Such an occurrence not only explains the quadrivalent arrangement of certain chromosomes in Tango, it also provides one possible explanation for the seedless characteristic of Tango fruit.

According to Ye et al (2009) embryonic degeneration in mandarin cultivars resulted in the production of seedless fruit. Male sterility due to chromosomal rearrangements that cause abnormal meiosis in microsporocytes is a main contributor to embryonic degeneration, and thus is a key component in the seedlessness of a fruit. Because the gametic cells that were studied in this project came directly from anthers (i.e exclusively male tissue) extracted from immature flower buds collected from Tango and W. Murcott plants, the results suggest that male sterility is likely responsible for the abnormal chromosomal arrangements seen in Tango and thus could also be responsible for its seedless phenotype.

Moreover, other oddities found in Tango gametic cells such as lagging chromosomes were seen in anaphase of meiosis I. According to Sheidai et al (2006) meiotic abnormalities such as the formation of a laggard chromosome may cause pollen sterility. The results that were found make a strong case for male sterility in the Tango mandarin, it is entirely possible that there is a direct relationship between indicators for male sterility found in Tango and the lack of seeds present in Tango fruit. This is especially probable due to normal chromosomal alignment found in metaphase/anaphase stages of meiosis in W. Murcott gametic cells; no significant deviances from normal chromosomal alignment and separation were observed in W. Murcott cells, which, according to Ye et al (2009), suggests that the development of microsporocytes in W. Murcott is normal. W. Murcott is capable of producing seedy fruit when cross-pollinated with a compatible variety of *citrus*, according to Wright et al (2007). While W. Murcott has a self-incompatibility system that prevents selfing and results in seedless fruit production when grown in isolation (Chao *et al* 2005), it has been known from experience in California that W. Murcott and Clementine mandarins are compatible with each other and seeds can be produced if they are planted in close proximity with one another due to cross-pollination (Chao *et al* 2005). This suggests that W. Murcott pollen is viable, and thus microsporocyte development is normal. Our results showing standard chromosomal alignment and separation during metaphase/anaphase stages of meiosis in W. Murcott gametic cells support these findings and also demonstrate a stark contrast with the chromosomal arrangements found in Tango gametic cells.

The methods of chromosomal analysis described here enable certain characteristics of chromosomes, such as the presence of bivalent and quadrivalent chromosomal formations, to be depicted when analyzing chromosomal misalignments. As technology further advances, there may one day exist equipment sensitive enough to detect and identify the slightest banding patterns given off by fluorescent dyes in FISH. Perhaps when this technology becomes available, it can be utilized in future studies which could identify each individual chromosome that is involved in such bivalent and quadrivalent formations.

References

Badaeva E.D., Dedkova O.S., Gay G., Pukhalskyi V.A., Zelenin A.V., Bernard S., and Bernard M., 2007. Chromosomal rearrangements in wheat: their types and distribution. Genome 50: 907-926.

Brasileiro-Vidal A.C., dos Santos-Serejo J.A., dos S. Soares Filho W. and Guerra M., 2006. A simple chromosomal marker can reliably distinguishes *Poncirus* from *Citrus* species. Genetica 129: 273–279.

Carvalho R., Soares Filho W.S., Brasileiro-Vidala A.C. and Guerraa M., 2005. The relationships among lemons, limes and citron: a chromosomal comparison. Cytogenet Genome Res 109: 276–282.

Chao C.T., 2005. Pollination study of mandarins and the effect of seediness and fruit size: implications for seedless mandarin production. HortScience 40: 362-365.

Curtis M.J., Belcram, K., Bollmann S.R., Tominey C.M., HoVman P.D., Mercier R., and Hays J.B., 2008. Reciprocal chromosome translocation associated with TDNA- insertion mutation in *Arabidopsis*: genetic and cytological analyses of consequences for gametophyte development and for construction of doubly mutant lines. Planta 229:731–745.

de Vozzi M.S., Santos S.A., Pereira C.S., Cuzzi J.F., Laureano L.A., Franco Jr. J.G. and Martelli L., 2009. Meiotic segregation and interchromosomal effect in the sperm of a double translocation carrier: a case report. Molecular Cytogenetics 2: 1-6.

Evans H.J., and Bigger T.R.L., 1961. Chromatid aberrations induced by gamma irradiation 11. Non randomness in the distribution of chromatid aberrations in relation to chromosome length in *Vicia faba* root-tip cells. Genetics 46: 277-289.

FAOSTAT (2011). Agriculture data. Available at http://www.faostat.fao.org. Accessed: 18 October 2011.

Henry I.M., Dilkes B.P., Comai L., 2007. Genetic basis for dosage sensitivity in *Arabidopsis thaliana*. PLoS. Genet. 3:e70.

Hensz R.A., 1982. Bud mutations in citrus cultivars in Texas. p. 89-91. In: K. Matsumoto (ed.), Proc. Intern. Soc. Citriculture , Vol. 1, Okitsu, Japan.

Hernandez-Montoya V., Montes-Moran M.A., Elizalde-Gonzalez M.P., 2009. Study of the thermal degradation of citrus seeds. Biomass and Bioenergy 33: 1295-1299.

Heslop-Harrison P., Osuji J., Hull R. and Harper G., 1999. Fluorescent in situ hybridization of plant chromosomes: illuminating the Musa genome. INIBAP: 26-29.

Li W.B., Pria Jr.W.D., Lacava P.M., Qin X. and Hartung J.S., 2003. Presence of *Xylella fastidiosa* in sweet orange fruit and seeds and its transmission to seedlings. Bacteriology 93: 953-958.

Lu B.C., 2006. Karyotyping of *Neurospora crassa* using synaptonemal complex spreads of translocation quadrivalents. Genome 49: 612-618.

Mabberley D.J., 1997. A classification for edible *Citrus* (Rutaceae). Telopea 7: 167-172.

Masoudi-Nejad A., S. Nasuda, R.A. McIntosh and T.R Endo. 2002. Transfer of rye chromosome segments to wheat by a gametocidal system. Chromosome Res. 10: 349-357

Marcucci M.C., Ragazzini D., Sansavini S., 1984. The effects of gamma and laser rays on the functioning of apple pollen in pollination and mentor pollen experiments. J. Hort. Sci. 59: 57-61.

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

Nasr-Esfahani M.H., Aboutorabi R., Shahnaz Razavi S., 2009. Credibility of chromomycin A3 staining in prediction of fertility. Intern. J. Fertility and Sterility 3: 5-10.

Raza H., Khan M.M., and Khan A.A., 2003. Seedlessness in citrus. Int. J. Agri. Biol. 5: 388-391.

Rickards G.K., 1983. Orientation behavior of chromosome multiples of interchange (reciprocal translocation) heterozygotes. Ann. Rev. Genet. 17: 443-498.

Roose M.L., Schwarzacher T., and Heslop-Harrison J.S., 1998. The chromosomes of *Citrus* and *Poncirus* species and hybrids: identification of characteristic chromosomes and physical mapping of rDNA loci using in situ hybridization and fluorochrome banding. J. Hered. 89: 83-86.

Roose M.L. and Williams T.E., 2007. Mandarin tree named ‘Tango’. US Plant Patent PP17,863.

Sanamyan M.F. and Rakhmatullina E.M., 2003. Cytogenetic analysis of translocations in cotton. Plant Breeding 122: 511—516.

Sax K., 1938. Chromosome aberrations induced by X-rays. Genetics 3: 494-516.

Shamel A.D., 1943. Bud variation and bud selection. In: H. J. Webber and L. D. Batchelor (eds.) The Citrus Industry, Vol 1. p. 915-952. Univ. of California Press, Berkeley.

Sheidai M., Golestanipoor A., and Jorjani E., 2006. Chromosome pairing and heterozygote translocation in Oltan cotton cultivar and its crossing progenies. Iranian Journal of Science & Technology 30 (A1): 103-109.

Viccini L.F. and de Carvalho C.R., 2002. Meiotic chromosomal variation resulting from irradiation of pollen in maize. J. Appl. Genet. 43: 463-469.

Wright G.C., 2007. Pollination of W. Murcott afourer mandarins. Citrus Research Report Dept. of Plant Sciences U. of A. Yuma Mesa Agri. Center 153:12-13.

Ye W., Qin Y., Ye Z., Teixeira da Silva J.A., Zhang L., Wu X., Lin S. and Hu G., 2009. Seedless mechanism of a new mandarin cultivar ‘Wuzishatangju’ (*Citrus reticulata* Blanco). Plant Science 177: 19–27.

Table 2.1: Number of Slides Viewed and Acceptable Chromosome Spreads Used in Data Analysis

Cultivar	Number of Slides Viewed	Average Number of Complete Chromosome Spreads Per Slide	% of Chromosome Spreads in Metaphase	% of Chromosome Spreads in Anaphase	% of Chromosome Spreads in Other Phase
W. Murcott	32	1.53	53.1	42.9	4.1
Tango	29	1.97	47.4	35.1	17.5

Table 2.2: Number of Images Successfully Taken With Leica SP2 confocal

Cultivar	Metaphase	Anaphase	Other Phase	Total
W. Murcott	12	7	1	20
Tango	31	11	7	49

Table 2.3: W. Murcott Chromosomal Spread Analysis (Percentages of Total Number of Complete Spreads Analyzed)

Meiotic Stage	NA	NS	MA	PMA	Q	A	LC	O
Metaphase	60	0	0	0	0	0	0	0
Anaphase	0	30	0	0	0	0	0	0
Other	0	0	0	0	0	0	0	10

NA=Normal Alignment; NS=Normal Segregation; MA=Misalignment; PMA=Partial Misalignment; Q=Quadrivalent A=Aneuploidy; LC=Lagging Chromosome; O=Other

Table 2.4: Tango Chromosomal Spread Analysis (Percentages of Total Number of Complete Spreads Analyzed)

Meiotic Stage	NA	NS	MA	PMA	Q	A	LC	O
Metaphase	0	0	23	7	9	0	3	0
Anaphase	0	0	0	0	2	11	2	0
Other	0	0	0	0	7	0	0	7

NA=Normal Alignment; NS=Normal Segregation; MA=Misalignment; PMA=Partial Misalignment; Q=Quadrivalent A=Aneuploidy; LC=Lagging Chromosome; O=Other

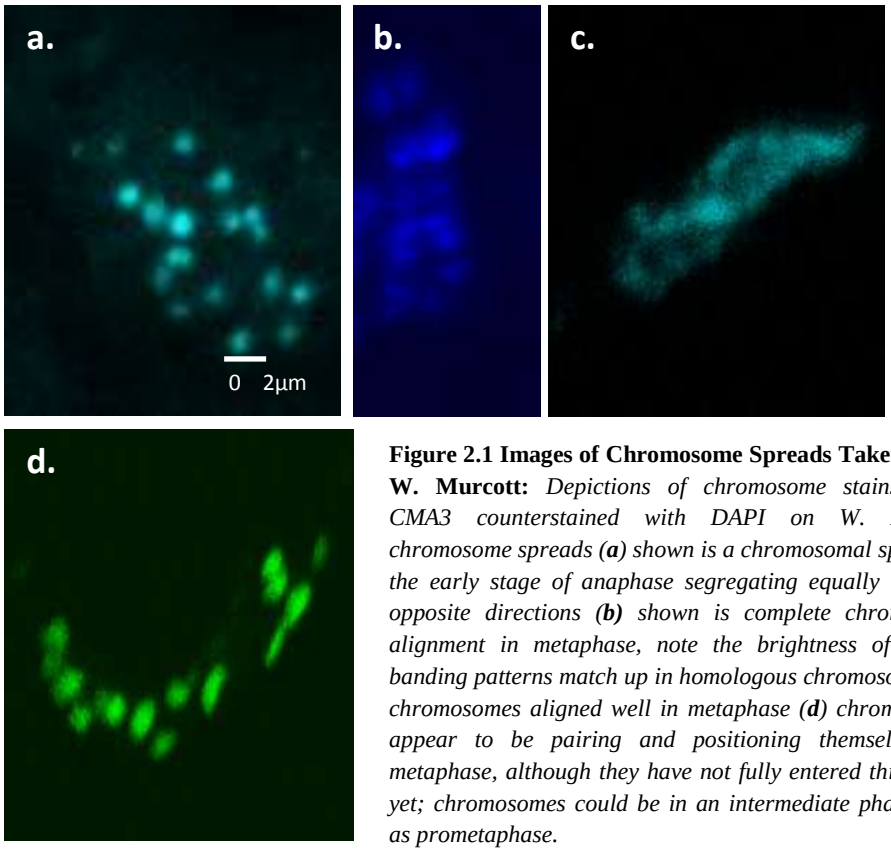


Figure 2.1 Images of Chromosome Spreads Taken From W. Murcott: *Depictions of chromosome stains using CMA3 counterstained with DAPI on W. Murcott chromosome spreads* ***(a)*** *shown is a chromosomal spread in the early stage of anaphase segregating equally towards opposite directions* ***(b)*** *shown is complete chromosome alignment in metaphase, note the brightness of CMA3 banding patterns match up in homologous chromosomes* ***(c)*** *chromosomes aligned well in metaphase* ***(d)*** *chromosomes appear to be pairing and positioning themselves for metaphase, although they have not fully entered this phase yet; chromosomes could be in an intermediate phase such as prometaphase.*

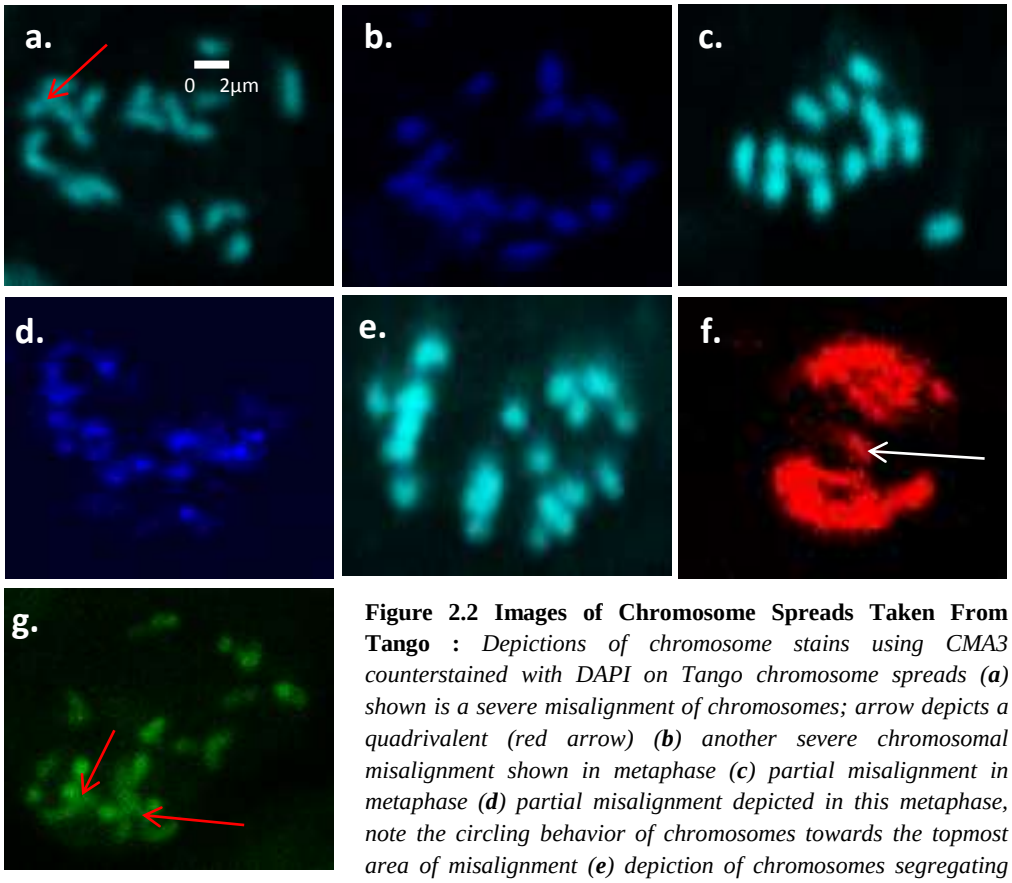


Figure 2.2 Images of Chromosome Spreads Taken From Tango : *Depictions of chromosome stains using CMA3 counterstained with DAPI on Tango chromosome spreads (**a**) shown is a severe misalignment of chromosomes; arrow depicts a quadrivalent (red arrow) (**b**) another severe chromosomal misalignment shown in metaphase (**c**) partial misalignment in metaphase (**d**) partial misalignment depicted in this metaphase, note the circling behavior of chromosomes towards the topmost area of misalignment (**e**) depiction of chromosomes segregating unequally in anaphase, thus creating aneuploidy, and (**f**) another abnormal chromosomal segregation in anaphase, note the lagging chromosome as shown by the arrow (**g**) quadrivalents (red arrows) appear to be responsible for the chaotic array of chromosomes pictured in this image.*

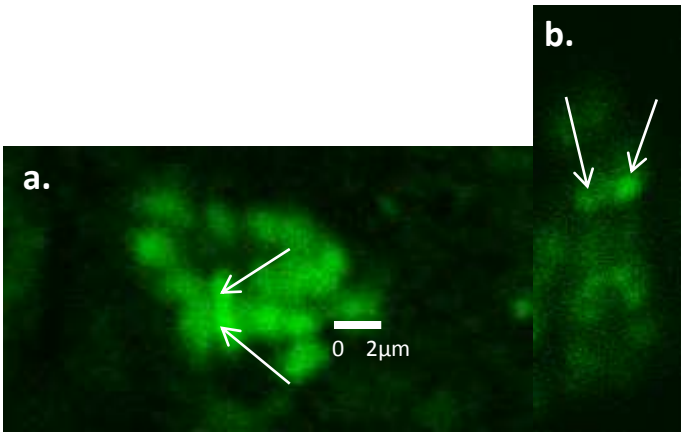


Figure 2.3 CMA3 Banding Patterns: *CMA3 brightness as depicted in chromosomal analysis of Tango (**a**) and W. Murcott (**b**) help to identify chromosomal homology but are not distinct enough to identify chromosomes individually. Zooming in on chromosomal images only further blurs the image as more graininess is introduced.*

Chapter 3

Analysis and Comparison of Copy Number of CIRE1, a *Ty1-copia* Retrotransposon, in Diverse *Citrus* Genomes

Abstract

Rico-Cabanas and Martinez-Izquierdo (2006) reported the first complete retrotransposon identified in *Citrus* (named CIRE1). Their report identified CIRE1 as an active, *copia* retrotransposon, present in the *C. sinensis* genome at around 2,200 full-length copies, contributing to 2.9% of the *C. sinensis* genome. The purpose of our study was to analyze the CIRE1 retrotransposon in two varieties of mandarin: W. Murcott, a seedy mandarin, and Tango, a seedless mandarin and gamma-irradiated offshoot of W. Murcott. The copy number of CIRE1 was compared between these two mandarin varieties, the hypothesis being that CIRE1 was activated in response to gamma radiation treatment received by W. Murcott budwood, which may have altered gene expression in the bud that gave rise to the seedless Tango variety. In order to estimate the copy number of CIRE1, quantitative PCR (qPCR) was performed using primers that targeted the conserved right-LTR region of CIRE1. Template DNA was extracted from young leaf tissue of the following *Citrus* varieties: Parent Washington Navel, Hart's Tardiff Valencia, W. Murcott, Tango, and Abhayapuri Lime. Additionally, Southern blot quantitation using probes targeting the right-LTR region, *int* gene, and *rt* gene of CIRE1 was performed to provide supporting

evidence. Additional supporting evidence was also provided by BLASTing the full-length right-LTR, *int*, and *rt* CIRE1 sequences against the *C. sinensis* and *C. clementina* genomes, as well as the W. Murcott/Tango assembly. From the qPCR analysis, 12 copies of the right-LTR fragment were found in Parent Washington Navel, 8 copies of the right-LTR fragment were found in Hart's Tardiff Valencia, 14 copies of the right-LTR fragment were found in W. Murcott, 14 copies of the right-LTR fragment were found in Tango, and 1 copy of the right-LTR fragment was found in Abhayapuri Lime. Southern blot analysis gave results that were difficult to interpret. The probes bound to a single region which was in the same location for W. Murcott, Tango, and Abhayapuri Lime digested genomic DNAs, though the probe signals were weak. The number of significant hits from BLAST against the *C. sinensis* and *C. clementina* draft genome assemblies ranged from 19 to 589, with 4 to 71 perfect matches. The number of significant hits generated from BLAST against the W. Murcott/Tango assembly ranged from 194 to 767, with 0 to 25 perfect matches. There is no significant difference in CIRE1 copy number between W. Murcott and Tango, meaning that the retrotransposon was likely not activated by gamma irradiation. In addition, the number of CIRE1 copies in the *Citrus* genome was far less than previously reported. The data presented in this study provides a foundation for further experimental studies to determine the correct number of CIRE1 retrotransposons that exist within the *Citrus* genome.

Introduction

Mandarin (*Citrus reticulata* Blanco) as well as pummelo (*Citrus maxima* Burm. Merrill) and citron (*C. medica* L.) have long been considered to be the three true *Citrus* species (Campos *et al* 2005). Mandarin is placed as the second most important citrus plant in the world (Campos *et al* 2005). The mandarin group is composed of a number of species and interspecific hybrids, making them the most phenotypically heterogeneous of the three true species of *Citrus* (Campos *et al* 2005). There have been multiple approaches to analyzing the genetic diversity of *Citrus* cultivars. These have included Restriction Fragment Length Polymorphism (RFLP) marker analysis to assay genetic similarity among *Citrus* species (Federici *et al* 2000, Nicolosi *et al* 2000), microsatellites (Barkley *et al* 2006, Froelicher *et al* 2008), Random Amplified Polymorphic DNA analysis (RAPD) (Sugarawa and Oowada 1995), Inter-Simple Sequence Repeat (ISSR) analysis, which has been used in *Citrus* identification (Fang *et al* 1998), and Inter-Retrotransposon Amplified Polymorphism (IRAPs) which has primarily been used in Clementina mandarin identification (Campos *et al* 2005, De Felice *et al* 2008, Gupta and Rustgi 2004). IRAP products are produced from two neighboring retrotransposons using outward-facing primers, which target conserved end-regions of the retrotransposon. These primers, when used in a polymerase chain reaction, amplify regions between retrotransposons, producing marker bands (Kalendar and Schulman 2007).

Retrotransposons make good candidates for DNA markers because their integration creates new junctions between neighboring retrotransposons and their conserved ends (Kalendar and Schulman 2007). Retrotransposon characteristics of integration activity, endurance, dispersion, preserved structure and sequence motifs, and high copy number imply that they are highly suitable candidates on which to build molecular marker systems (Kalendar and Schulman 2007). The majority of the retrotransposon marker methods utilize two fundamental features. Firstly, retrotransposons produce large insertions as a result of their transpositional activity and secondly they contain conserved domains from which PCR primers can be designed (Kalendar *et al* 2011).

Retrotransposons are mobile genetic sequences, which self-replicate through the reverse transcription of RNA intermediates (Takeda *et al* 1998, Kumar and Bennetzen 1999). Retrotransposons move to new areas in the genome via an RNA intermediate that is transcribed into DNA by a retrotransposon-encoded reverse transcriptase/RNaseH enzymes before being subsequently inserted into the genome (Kumar and Bennetzen 1999). This replicative method of transposition can swiftly increase the copy numbers of retroelements and therefore can significantly increase plant genome size (Kumar and Bennetzen 1999). The retrotransposon life cycle results in genome expansion and because of this property their activity is stringently controlled by the host (Manetti *et al* 2007). In addition to rounds of amplification, the size of the host genome can be decreased by incidents of homologous recombination within an element or between elements, typically increasing the copy number of SOLO-LTR sequences (Manetti *et al* 2007). Deviations in genome size and the C-value paradox can be ascribed to differences in the amount of

transposable elements, retrotransposons in particular, amongst the genomes of different species of plants (Pereira *et al* 2004).

Retrotransposons are universal in plants and play a vital role in plant gene and genome evolution (Kumar and Bennetzen 1999). Usually, retrotransposons make up 50% or more of nuclear DNA content, a condition that may occur within a few million years (Kumar and Bennetzen 1999). Some plant genomes have inflated due to their capacity to tolerate considerable amplification of retrotransposons. In smaller genomes, these elements are observed in lower copy numbers, are non-arbitrarily situated within the genome and are primarily restricted to heterochromatic areas (Pereira *et al* 2004). It is theorized that the non-arbitrary genomic dispersion of retrotransposons in some species reveals selection against hosts which carry detrimental retrotransposon insertions in particular genomic regions (Pereira *et al* 2004). Within smaller genomes, grouping of retrotransposon insertions in compact heterochromatin, which has decreased frequencies of recombination as well as low gene density and levels of transcription, is presumably consistent with the idea of negative selection and of passive buildup of retrotransposons where their incorporations would be less deleterious (Pereira *et al* 2004). As another possibility to selection, one hypothesis to rationalize this grouping of retrotransposons concerns preferential insertion, or even positive selection for their preservation within heterochromatin (Pereira *et al* 2004).

Long terminal repeat (LTR) retrotransposons tend to be the dominant retrotransposon class in plants. The first plant LTR retrotransposons described were mostly regarded as by-products of defective element characterization, or were discovered

by chance while seeking out polymorphisms (Vitte *et al* 2005). Later, efforts to identify genes of interest, through genetic marker analysis and the assembly of small insert genomic libraries, also led to the fortuitous characterization of LTR retrotransposons (Vitte *et al* 2005). More recently, analysis of large insert libraries, and genome sequencing projects, have resulted in numerous large plant genomic sequences and therefore facilitated the discovery of new LTR retrotransposons (Vitte *et al* 2005).

The LTR retrotransposons have LTRs and these elements can range from a few hundred bp to over 5 kb in size (Kumar and Bennetzen 1999). Like retroviruses, the LTR retrotransposon genome contains two sequentially conserved LTR regions (Pereira *et al* 2004). The LTRs do not code for any known proteins, but they do contain the promoters and terminators that facilitate the transcription of LTR retrotransposons (Kumar and Bennetzen 1999). They are sequences of DNA that flank an internal region (IR) that typically contains genes and other features necessary for self-directed retrotransposition (Pereira *et al* 2004).

LTR retrotransposons code for proteins denoted by three major genes: *gag*, *pol*, and *int* (Kumar and Bennetzen 1999). These genes and proteins are specified by a single mRNA molecule with the arrangement 5'—R–U5—PBS–coding region–PPT–U3—R–3'; where R, U5,PBS, PPT, and U3 denote repeated RNA, unique 5' RNA, primer-binding site, polypurine tract, and unique 3' RNA, respectively (Kumar and Bennetzen 1999). Transcription is initiated at the 5' end of R in the 5' LTR and is terminated in the 3' end of R in the 3' LTR (Kumar and Bennetzen 1999). The *gag* gene codes for proteins required for the development and packaging of retrotransposon RNA and proteins into an

appropriate arrangement for incorporation into the genome (Kumar and Bennetzen 1999). The *pol* gene encodes reverse transcriptase and RNase H enzyme activities that are required for replication/transposition of the retrotransposon, whereas *int* encodes for the integrase enzyme that allows the DNA form of the retrotransposon to insert at a new chromosomal location (Kumar and Bennetzen 1999).

Retrotransposons, much like DNA transposable elements, can induce mutations when inserting near or within genes (Kumar and Bennetzen 1999). Furthermore, mutations caused by retrotransposon insertion are fairly stable; this is because the sequence at the insertion site is preserved due to transposition via replication (Kumar and Bennetzen 1999). Transposons are capable of gene inactivation through integration or by triggering methylation in the area where they are located (Biswas *et al* 2010). They may also be involved in agronomic variation (such as maturation date, flesh color) through the expansion of variety in alleles or through interference in gene expression (Biswas *et al* 2010).

Retrotransposons are grouped into two large classes; based on the existence or absence of long terminal repeats (LTR) at both ends (Takeda *et al* 1998) The basic structure of LTR retrotransposons resembles that of proviruses or retroviruses; both contain internal genes which code for Pol and Gag proteins that are necessary for reverse transcription and integration within the host genome (Takeda *et al* 1998). The gene that codes for the Env protein and is necessary for retroviruses to propagate from a host cell to other cells is not present within LTR-retrotransposons (Takeda *et al* 1998). LTR-retrotransposons are further divided, by the order of internal domains and their sequence

homology, into two groups, the Ty1-*copia* and Ty3-*gypsy* groups (Takeda *et al* 1998). Ty1-*copia* retrotransposons exist throughout the plant kingdom, in species varying from single-cell algae to bryophytes, gymnosperms, and angiosperms (Kumar and Bennetzen 1999). Ty3-*gypsy* retrotransposons also are known to be broadly dispersed in the plant kingdom, including both gymnosperms and angiosperms (Kumar and Bennetzen 1999). In addition, both these groups of retrotransposons are frequently found in high copy number, as many as a few million copies per haploid nucleus, in plants having large genomes (Kumar and Bennetzen 1999).

Plant retrotransposons are structurally and functionally analogous to the retrotransposons and retroviruses that belong in other eukaryotic organisms, however, there are significant dissimilarities in the genomic composition of retrotransposons in plants compared to some other eukaryotes. These differences include their often-high copy numbers and their chromosomal distribution patterns (Kumar and Bennetzen 1999). PCR-based analysis has demonstrated that Ty1-*copia* retrotransposons are pervasive in plant genomes (Takeda *et al* 1998). However, the majority of plant retrotransposons have been inactive, and many of them appear to have acquired their inactivity due to their defective structures (Takeda *et al* 1998). In order to safeguard the continued existence of their host, and thus their own survival, host genomes have evolved stringent regulators of retrotransposon activity (Grandbastien *et al* 2005). This regulation can take place at various stages during the process of retrotransposition, one of the most significant being transcriptional regulators that control the manufacturing of RNA, without which, retrotransposons are unable to generate daughter copies (Grandbastien *et al* 2005).

Furthermore, evolutionary controls and selection that affect groups of hypothetically active copies influence both the location and the effect of retrotransposons on their host genomes (Grandbastien *et al* 2005). Control mechanisms have been analyzed in detail for the Tnt1 retrotransposons discovered in species of the Solanaceae family (Grandbastien *et al* 2005). The novel Tnt1A element of tobacco was isolated after having transposed into a target gene (Grandbastien *et al* 1997). Tnt1A expression is strongly regulated; it is not transcribed in unstressed tobacco plants with the exception of low amounts of transcription in the roots (Grandbastien *et al* 1997). However, Tnt1A becomes expressed at high levels when induced by stress conditions; wounding, biotic elicitors and pathogen attacks are a few examples of these stress conditions (Grandbastien *et al* 1997). Some LTR retrotransposons that are largely inactive during development can be transcriptionally activated and transpose under conditions of biotic and abiotic stress (Feschotte *et al* 2002). After stress-induced transcription, the genomic copy number of the rice LTR retrotransposon *Tos17* increased from 2 to more than 30 copies in some lines, as determined through Southern blotting (Feschotte *et al* 2002).

Tto1 and *Tto2* of tobacco and *Tos17* of rice, identified as transcriptionally active retrotransposons, were shown to be mobile in cultured cells (Takeda *et al* 1998). Transposition of *Tto1* and *Tos17* was associated with an increase in RNA levels, signifying that retrotransposon transposition is primarily regulated at the transcriptional level (Takeda *et al* 1998). The tobacco Tnt1 retrotransposon, the first active retrotransposon discovered in plants, was proposed to have been activated during protoplast preparation because Tnt1 transcription was stimulated during protoplast

preparation (Takeda *et al* 1998). It has been demonstrated that activation of transcription of Tnt1 is induced by microbial elicitors of plant defenses (Takeda *et al* 1998). Transcription activation of *Tto1* and Tnt1 may also be initiated through virus infection (Takeda *et al* 1998). These results suggest that additional biotic or abiotic stresses could activate various plant retrotransposons, which are dormant under normal growth conditions (Takeda *et al* 1998). Transcription of retrotransposons is generally controlled by regulatory regions in the 5' LTR as well as neighboring untranslated regions (Takeda *et al* 1998).

In recent years, Rico-Cabanas and Martinez-Izquierdo reported the first complete retrotransposon identified in *Citrus* (named CIRE1) and showed that it has all the features of a typical *copia* retrotransposon (Rico-Cabanas and Martinez-Izquierdo 2006). The CIRE1 retrotransposon was fully sequenced by this team. Using a slot blot hybridization method, it was found that CIRE1 was present in the *C. sinensis* genome at around 2,200 full-length copies, contributing to 2.9% of the *C. sinensis* genome (Rico-Cabanas and Martinez-Izquierdo 2006).

Because retrotransposon activation, proliferation, and subsequent insertion can sometimes alter gene expression, it was decided to compare and contrast the copy number of a target retrotransposon between two varieties of mandarin, W. Murcott and Tango. As outlined below, Tango was selected from a population of trees grown from gamma irradiated buds of W. Murcott. In theory, a retrotransposon could have been activated in response to the gamma radiation treatment received by W. Murcott budwood. This may

have altered the gene expression of the bud that gave rise to the seedless Tango variety. CIRE1 was selected because it is documented as a sequence that is both active and exists in high copy number in the citrus genome.

W. Murcott is a predecessor to Tango; in 1995, budwood cuttings were taken from a W. Murcott tree and were exposed to gamma radiation. Individual buds from the cuttings were then grafted onto rootstocks and the resulting trees were observed over time to determine the genetic effects of the radiation exposure. While many of the trials did not produce desirable physical attributes (many plants died prematurely or did not grow properly) one tree bore commercially desirable fruit (easy to peel, sweet tasting, seedless mandarins). Its fruit resembled that of W. Murcott in nearly every aspect, with the exception of its capability to maintain seedlessness. While W. Murcott fruit is also sweet, juicy, and easy to peel, it is seedy when cross-pollinated by other citrus varieties (oranges, tangelos, lemons, etc.). Tango, on the other hand, maintains a relatively seedless state (less than 0.2 seeds per fruit, on average) even when cross-pollinated with other citrus varieties (Roose and Williams 2007). If a retrotransposon, such as CIRE1, were activated by the stress-inducing effects of gamma radiation, and a copy inserted near or within genes crucial to the fertility of Tango, this mutation event could account for the seedless characteristic found in Tango. If the CIRE1 retrotransposon replicated during genesis of Tango, one would expect to find a higher copy number of CIRE1 in the Tango genome when compared to the number of CIRE1 copies found in the W. Murcott genome.

Quantitative PCR (qPCR) is the most accurate, precise, and efficient method currently available for DNA quantitation. It is an amplification method that detects and measures the accumulation of DNA that binds fluorescent dyes as the reaction progresses. Therefore, this method was chosen to estimate the copy number of the CIRE1 retrotransposon sequence in Tango, W. Murcott, and other citrus cultivars. Southern blot quantitation and BLAST against the *C. sinensis* and *C. clementina* genomes provided additional supporting evidence of the estimated copy number.

A more reliable estimation of the copy number of this sequence will help progress our understanding of the composition of the citrus genome. In this experiment, several varieties of citrus plants were chosen for observation, these were: Parent Washington Navel (a *C. sinensis*), Hart's Tardiff Valencia (variety studied by the group who discovered CIRE1), W. Murcott (predecessor to Tango), Tango (product of the irradiation of W. Murcott budwood), and Abhayapuri Lime (used as an outgroup).

Materials and Methods

Primer Set Design

Primer sets were designed from the CIRE1 retrotransposon nucleotide sequence available in the EMBL database under the accession number AM040263. The right LTR-sequence of the CIRE1 retrotransposon was input into the Primer3 primer design online program (accessed via http://frodo.wi.mit.edu/primer3/), using the default parameters outlined in the program for forward and reverse primer selection. The amplicon size was set as 100-300 bp. The primers selected and used to amplify the right LTR-region of interest are as follows: forward primer (5'-ATAACGAAGCCCAGCACACT-3') reverse primer (5'-CAAAAGACACAGACCAGCAAA-3'). This primer set is referred to in the experimental analysis as 2F/2R. These primers are located at bases 4716-4735 and 4908-4928 of the CIRE1 sequence. The primer set used for normalization of the data was obtained from the USDA Repository located in Riverside, California, and it was used to amplify a region of a gene identified as coding Phospholipase D. The sequences of these primers are as follows: forward primer (5'-ATCCTCTGCAAGTAGATGC-3') reverse primer (5'-CAACATATCAGGAAGACTCATG-3').

Tissue Collection/DNA extraction

All DNA samples were extracted from young leaf tissue of trees growing at the Citrus Research Center at the Univerisity of California, Riverside. DNA was extracted from a W. Murcott tree located in field 12D row 3 tree 4, the original or "mother" Tango tree, located in field 15F row 2 tree 10, a Parent Washington navel orange tree located in field

12B row 6 tree 4, a Hart's Tardiff Valencia orange tree located in field 12B row 7 tree 9, and an Abhayapuri lime tree located in field 18B row 25 tree 2. Leaves were washed of dirt and residue by rinsing in cold tap water and blotted with paper towels. After washing, leaves were ground into a fine powder with the use of liquid nitrogen and a mortar and pestle. The pulverized tissue was kept frozen on dry ice, and was later transferred to a -80°C freezer until ready to be used. To extract the DNA, ~100 mg of frozen tissue (wet weight) was extracted using a DNeasy Plant Mini Kit (Qiagen) following the manufacturer's instructions. All steps were followed precisely with the exception of step 19; rather than pipetting an additional 100 µl Buffer AE onto the DNeasy membrane, the 100 µl of Buffer AE that had already passed through the membrane was taken up by a pipette and was again pipetted onto the Dneasy membrane and centrifuged into a microcentrifuge collecting tube. This was done in order to enhance the concentration of extracted DNA. Once the DNA was extracted, DNA concentration was measured with a Nanodrop 3300 spectrophotometer. This DNA concentration in ng/µl was used to estimate the amount of DNA when making dilutions used in the qPCR reaction.

For qPCR, SYBR Green was chosen as the fluorescent label to use in the reaction. The amounts for each component of the mix used for the qPCR reaction were as follows: 12.5 µl SYBR Green/Rox PCR master mix (SABiosciences, 2X concentration), 0.5 µl forward primer (10 µM), 0.5 µl reverse primer (10 µM), 9.5 µl ddH_2O, 2 µl diluted DNA. DNA was diluted 1:1000 from its stock concentration in ddH_2O. DNAs were further diluted by 1:2, 1:4, 1:8, and 1:16 dilutions in ddH_2O. A 96-well plate was organized to include all citrus varieties (Parent Washington, Valencia, W. Murcott, Tango mothertree, and

Abhayapuri Lime) and their dilutions with a primer set amplifying the right LTR region of the CIRE1 retrotransposon.

Normalization

Data normalization in real-time quantitative-PCR is a further major step in DNA fragment quantification analysis. The reliability of any relative RT-qPCR experiment can be improved by including an invariant endogenous control (reference gene) in the assay to correct for sample to sample variations in RT-qPCR efficiency and errors in sample quantification. For this analysis, a gene that encodes the enzyme Phospholipase D available in GenBank under the EST accession number CX074707, was selected; this gene is present in a single copy in the *Citrus clementina* genome sequence (www.phyotozome.net) and is likely present in the same copy number for all *Citrus* varieties studied. Therefore it aids in selection of dilutions from each variety that contain a comparable amount of DNA. Primer sets used in qPCR amplification of Phospholipase D are (584/585).

Standard Curve

A standard curve was constructed from DNA of known concentration. This curve was then used as a reference standard for extrapolating quantitative information for DNA targets of unknown concentrations. To accomplish this, the right LTR region of CIRE1 retrotransposon was PCR cloned into a pCR®4-TOPO® plasmid. The concentration of the plasmid containing the DNA fragment of interest was measured with the Nanodrop spectrophotometer, and then ddH_2O was used to make a series of ten-fold dilutions of the concentrated stock; dilutions began at 10^{-1} and continued onward by a factor of ten,

concluding with a 10^{-8} dilution. Next, qPCR reactions were run on this dilution series. These initial experiments indicated that a standard curve was optimal with DNAs from 10^{-6} to 10^{-8} dilutions. DNAs were diluted by a factor of 0.5. Therefore the series started with 10^{-6} 1:2 and ended with 10^{-6} 1:64, concluding with a 10^{-8} dilution. The equation fit to this standard curve was: Y= -3.378*log(X) + 17.05, where X is equal to the initial template quantity in the Standard wells and Y is equal to the threshold cycle.

Cloning

Clones used for the standard curve were created using a TOPO®TA Cloning®Kit for Sequencing from Invitrogen. A PCR reaction using W. Murcott as a DNA template and primer sets targeting the right LTR region of CIRE1 (2F/2R) was set up according to the vendor's instructions. The PCR product was cloned using a TOPO® Cloning reaction, which was performed according to the manual's instructions. pCR®4-TOPO® construct was transformed via heat shock into the chemical competent *E. coli* cells provided with the kit according to the manual's instructions.

qPCR Protocol

The thermocycler used for quantitative polymerase chain reaction (qPCR) is a Stratagene MX 3005P, located at the USDA Citrus Repository; the amplification protocol used was a standard 40 cycle, two-step qPCR reaction completed in 90 minutes. Temperature settings and cycling parameters were as follows: initial temperature of 50°C for two minutes, followed by 95°C for ten minutes, followed by repeated, denaturation, annealing, and elongation temperatures of 95°C for 15 seconds, 50°C for 30 seconds, and

72°C for 30 seconds, respectively (repeated for a total of 40 cycles). Amplification was concluded with a single dissociation step.

Primer Set Design for Genomic DNAs

Primer sets were designed targeting three regions of the CIRE1 LTR retrotransposon: the reverse transcriptase (***RT***) gene, the integrase (***INT***) gene, and the right long-terminal repeat (**LTR**) region. Primers for these three target regions were designed using the Primer3 online program in combination with primer sequences used by Rico-Cabanas et al (2005). Primers used by Rico-Cabanas et al were analyzed by the Primer3 program. Irregularities, such as primer dimerization or extreme differences in melting temperature (Tm) points, were examined. Primer sequences were slightly altered as necessary to avoid predicted problems. As a result, the primer sequences used were similar, but not identical, to those used by Rico-Cabanas et al. The primer sequences used in this study were:

-forward LTR primer (5'-GAGCTGTTGTGCAGATTG-3');

-reverse LTR primer (5'-CAGCAAGACTCACGAATC-3');

-forward INT primer (5'-GATAGAATCCAGCAACTG-3');

-reverse INT primer (5'-AGCTTTGAATTATCAGCATAC-3');

-forward RT primer (5'-GAAGGTGTAGATTTCAATG-3'); and

-reverse RT primer (5'-CTGACTTGTATTTGCTTGCA-3').

Enzyme Digestion of *Citrus* Genomic DNA

To fragment genomic DNA for Southern blot analysis, separate digestions were performed with two enzymes: EcoRI and HindIII (New England BioLabs). EcoRI was chosen because it is a frequent and efficient base-cutter, HindIII was chosen because of its lack of star activity and its insensitivity to DNA methylation. Another advantage of using different enzymes is that it results in different DNA fragment sizes and thereby improves estimates of the number of probe target sequences present. Digestions were performed as follows, for EcoRI digestion: 2-5 µg of genomic DNA was digested in 200-500µl total reaction volume (1µg DNA/100µl reaction volume). The final reaction contained 1X EcoRI-specific buffer (from 10X stock concentration, from NEB), 1X BSA (from 100X stock concentration), 2.5 mM spermidine, and 200-500 units of EcoRI enzyme (20 units/200 ng genomic DNA in reaction); ddH_20 was used to bring the reaction up to the correct volume. All components of the enzyme digestion reaction were added to a 2-ml microcentrifuge tube, with ddH_20 added first, DNA heated at 65°C for ten minutes added second to last, and EcoRI enzyme added lastly to the mixture. Digestions were then incubated at 37°C for two hours, digestion was stopped by incubating samples at 65°C for 20 minutes. For HindIII digestion: 2-5 µg of DNA was digested as described for EcoRI except that the final reaction contained 1X HindIII-specific buffer and 200-500 units of HindIII (from NEB). Digestions were incubated at 37°C overnight, digestion was stopped by incubating samples at 65°C for 20 minutes.

Both EcoRI and HindIII digestion reaction mixtures were vacuum concentrated to 20-40 µl for easier gel loading. To check for complete digestion of DNA, undigested genomic DNA of each sample was analyzed as a control. 0.9% agarose gels containing ethidium bromide (EtBr) at 0.1 µg/ml and 1X TBE buffer (0.089M Tris base, 0.089M boric acid, 0.002M EDTA, pH 8.0) were used to fractionate genomic DNAs, which were loaded at 150 ng/lane. HindIII-digested λ-DNA (NEB, 250 ng) and 100-bp DNA ladders (from Invitrogen, 550 ng) were loaded in the first two lanes. 6X-loading dye (0.25% bromophenol blue, 0.25% xylene cyanol FF, 40% (w/v) sucrose in water) was mixed with each sample at a concentration of 1µl dye for every 5 µl of sample. The gel was run at 90V for 40 min, followed by 40V for 30 min.

PCR Amplification of Regions of Interest

To generate PCR fragments pertaining to the regions of interest within the CIRE1 LTR retrotransposon, three separate PCR reaction mixtures were made. Each mixture contained the same components with the exception of the primer set which targeted reverse transcriptase (RT), integrase (INT), or 5' long-terminal repeat (LTR). Reactions included: 0.2 mM dNTPs, 2.0 mM $MgCl_2$, 0.05 U/µl GoTaq DNA polymerase (from NEB), 5 ng/µl W. Murcott genomic DNA, 0.05 µM forward primer, 0.05 µM reverse primer, and ddH_2O to 25 µl total reaction volume. The three reactions were placed in a heat-enabled lid thermocycler programmed with the following settings: initial 94°C step for five min; 45°C for two min; 72°C for three min; 94°C denaturation step one min; 45° annealing step two min; 72°C elongation step three min; repeat denaturation, annealing, and elongation steps 29 times; 94°C one min; 45°C two min; 72°C ten min; final 4°C

indefinite cooling step. After PCR amplification completion, concentration of each sample was measured with a Nanodrop Spectrophotometer, and PCR products analysed on an agarose gel to determine that a single PCR fragment of the expected size was present. A 0.9% agarose gel containing EtBr at a concentration of 0.01µg/ml EtBr of liquid agarose was prepared. 500 ng of 100-bp DNA ladder was loaded in the first lane, 200 ng of INT PCR fragment, 200 ng of RT PCR fragment, and 125 ng of LTR PCR fragment were loaded in three separate lanes. 1 µl 6X-loading dye was mixed with every 5 µl of sample that was loaded. The gel was run at 100V for 1.2 hours.

Southern Blot Procedure

For the genomic DNA blot, a 25 cm X 20 cm 0.9% agarose gel containing 0.01 µg/ml EtBr was prepared and loaded as follows: lane 1 loaded with 500 ng of HindIII-digested λ DNA ladder; lane 2 loaded with 550 ng of 100-bp DNA ladder; lane 10 loaded with ~2 µg HindIII-digested W. Murcott; lane 12 loaded with ~5 µg HindIII-digested Tango; lane 14 loaded with ~5 µg HindIII-digested Abhayapuri Lime; lane 16 loaded with ~4 µg EcoRI-digested W. Murcott; lane 18 loaded with ~5 µg EcoRI-digested Tango; lane 20 loaded with ~5 µg EcoRI-digested Abhayapuri Lime; lane 22 loaded with 50 ng of LTR PCR fragment; lane 23 loaded with 60 ng of INT PCR fragment; lane 24 loaded with 60 ng RT PCR fragment. PCR fragments were included for Southern blot transfer to serve as both positive and negative controls for the probe. 6X-loading dye was mixed with each sample before loading into the gel, at a concentration of 1 µl for every 5 µl of sample volume. The gel was run initially at 100V for 2 hours, followed by 40V for ~18 hours. The region of the gel containing HindIII and EcoRI digested samples, as well as the PCR

fragments, was excised and transferred to a positively charged nylon membrane (Nylon 66 plus from Hoefer Scientific). Further gel treatment and Southern blot transfer to positively charged nylon membrane was done following the protocol provided by Sambrook (Sambrook *et al* 1989). Prepared, neutralized nylon membrane was stored dry, in plastic wrap, at 4°C until needed for pre-hybridization.

Preparation and Labeling of Probe

Each PCR product to be used as a probe was labeled by random hexamer labeling (Sambrook *et al* 1989) as follows. A total of 25-50 ng of the PCR fragment chosen to be used as a probe was placed in a 1.5-ml screw cap microcentrifuge tube volumed to 10 µl with ddH_2O. The fragment was denatured by boiling for three min., then chilling on ice for at least 15 min. To this, two µl of hexanucleotide mix (from Roche), three µl of 0.5 mM dNTPs (consisting of equal amounts of dATP, dGTP, dTTP only) and one µl Klenow enzyme (5000 U/ml, from Roche) was added. The microcentrifuge tube was inverted a few times to mix components and was then spun down briefly. To this, four µl of ^{32}P dCTP (40 mCi) was added and the probe labeling reaction was allowed to incubate at room temperature overnight. To purify the probe, a Sephadex G50 column was prepared consisting of a Pasteur pipet plugged with glass wool and filled with Sephadex G50 equilibrated in 5X TE (50 mM Tris-HCl, pH 7.5; 5 mM EDTA, pH 8.0)/0.1% SDS; about 1 cm of space was left near the opening of the pipet to allow for the addition of 5X TE/0.1% SDS to keep the Sephadex from drying out. To the newly labeled, unpurified probe, 25 µl of nick translation stop dye (0.5% Bromophenol blue, 20% glycerol, 1.5% blue dextran, 60 mM $Na_2EDTA\text{-}2H_2O$, components dissolved in ddH_2O) was added and

mixed by slowly pipetting up and down, the entire probe/dye mixture was then loaded on top of the column. It was washed with 5X TE/0.1% SDS as the probe mixture progressed through the column. The first blue fraction (blue dextran) was collected in a new 1.5-ml screw cap microcentrifuge tube. The CPM of the probe was checked using a Geiger counter.

Hybridization of membrane

The nylon membrane was removed from plastic wrap and prehybridized in a hybridization bottle containing 20 ml Church's hybridization solution (http://rice.genomics.purdue.edu/docs/ OvergoHybProtocol.pdf) at 55°C for ≥ 1 hr using a hybridization oven with rotation. Labeled, purified probe was denatured by boiling three min., and placed on ice for at least 15 min. Denatured probe was added directly to the Church's hybridization solution in the hybridization bottle and was incubated at 55°C overnight with rotation. The membrane was rinsed with ~50 ml 2X SSC/0.2% SDS (100 ml 20X SSC, 20 ml 10% SDS in ddH_2O to 1 L) in the hybridization bottle for ~1 min before solution was disposed. Fresh, preheated ~50 ml of 2X SSC/0.2% SDS was added to the bottle and allowed to incubate at 55°C for 20 min. Solution was disposed and replaced with preheated 0.2X SSC/0.2% SDS (10 ml 20X SSC, 20 ml 10% SDS in ddH_2O to 1 L) solution and allowed to incubate at 55°C for 20 min. The solution was disposed and a second preheated 0.2X SSC/0.2% SDS wash was repeated. The membrane was wrapped in Saran wrap and was checked for probe activity using a Geiger counter. The level of CPM as measured by the counter was recorded.

Capturing signal of probe on film

The wrapped membrane was placed in a film cassette with intensifying screens and film (Premium Blue X-ray film, Biolane Products). The cassette was locked shut and was incubated at -80°C for 2-5 days, depending upon activity of the probe as measured by its CPM level. Film was then removed from the membrane in the dark room and developed.

BLAST Analysis

Basic Local Alignment Search Tool (BLAST) was performed using the complete, full-length 5044-bp CIRE1 retrotransposon sequence (accession number AM040263), the full-length 461-bp right-LTR CIRE1 sequence, the full-length 771-bp integrase gene sequence (found using Tal-3, a copia-like transposable element found by Voytas and Ausubel (1988) in *Arabidopsis thaliana*, as a template; amino acid sequence of translated integrase gene in Tal-3 was aligned with amino acid sequence of translated CIRE1 to identify the CIRE1 integrase gene), and the full-length 729-bp reverse transcriptase gene sequence (also found using Tal-3 as a template) as queries (*Appendix A*). The reference genome sequences that were used for this analysis were *C. clementina*, *C. sinensis* (both found in www.phytozome.net online database), and Tango/W.Murcott Abyss assembly (developed by Rebecca Sun and found on http://bioweb.ucr.edu/privateBlast/blast.html online database). For the phytozome database, the following parameters were selected: E threshold of -10, comparison matrix BLOSUM62 (default), word length of 11 (default), allow gapped alignment, unfiltered, and 500 alignments to show. For the Tango/W.Murcott database, all default settings were selected with the exception of the expected threshold value (0.0001), the number of output descriptions (500), and no

filters. The number of BLAST hits was recorded for each query, the most significant hits (e-value of 0, indicating a perfect match) were recorded separately.

Results

Regions of CIRE1 Targeted in qPCR Analysis

The region within the CIRE1 LTR retrotransposon targeted for quantitation in qPCR is a 247-bp sequence located in the right-LTR region of CIRE1 (*Figure 3.1*). This region was targeted and quantitated in the following varieties of *Citrus*: Parent Washington Navel, Hart's Tardiff Valencia, W. Murcott, Tango, and Abhayapuri Lime. To quantify the CIRE1 right-LTR fragment, a standard curve was generated using a dilution series of a clone of this fragment amplified from W. Murcott DNA (*Figure 3.2A*). The standard curve had an r^2 value of 0.965 and an efficiency of 97.7% (*Figure 3.2B*). A dissociation step was also performed to ensure that only the 247-bp LTR insert within the plasmid was subject to amplification and quantitation; peaks that are generated are representative of a single amplified product, signifying a single sequence was amplified (*Figure 3.2C*).

Normalization Curve Used for Dilution Selection

Normalization curves were generated for all five *Citrus* varieties using the housekeeping gene (Phospholipase D) as a reference gene. This gene is present in the same copy number for all *Citrus* varieties studied based on BLAST results which show the same number of hits against the *Citrus* genomes of interest (*Tables 3.1, 3.2*). The W.Murcott/Tango Abyss sequence assembly resulted in 3 perfect Phospholipase D BLAST hits, but this refers to a single hit on each heterozygous allele, so it is considered to be a single perfect match. Thus, only dilutions from each variety having similar cycle threshold (Ct) values to one another for this gene are later used for comparison for the

LTR DNA fragment of interest. The normalization curve (*Figure 3.3A*) is based on dilutions from the five *Citrus* varieties analyzed that have similar Ct values, as demonstrated further in the scatterplot of these values (*Figure 3.3B*).

Plotting Ct Values Against the Standard Curve

These same dilutions used in the normalization curve were used in the analysis of Ct values generated from the targeted 247 bp right-LTR region of CIRE1 (*Figure 3.4A*). Ct values generated from each *Citrus* variety dilution used in the comparison were plotted against the standard curve (*Figure 3.4B*) and copy number of the LTR fragment of interest was calculated for each *Citrus* variety studied using the standard curve to estimate the amount in picograms of LTR fragment within each dilution based on its Ct value, as well as the estimated genomic DNA concentration for each of the dilutions used. A dissociation curve generated from the qPCR amplification of the LTR region of interest (*Figure 3.4C*) shows a single distinct peak for each variety analyzed.

Calculations Used to Determine Copy Number of CIRE1

The calculations performed to determine the copy number of the right-LTR region in CIRE1 are outlined in Appendix B (*Appendix B*) and are based on the concentration of undiluted DNA as measured by Nanodrop spectrophotometer and the dilution used in the reaction. These calculated concentrations were converted to nucleotide base pairs and the copy number estimated using the percentage that the 247 bp right-LTR fragment comprises in the 3.82×10^8 bp citrus genome.

Copy Number of CIRE1 Found per Genome Analyzed

Calculated copy number of the right-LTR region of CIRE1 ranged from 1-12 copies per genome among the five cultivars tested. Two clones of the 247-bp right-LTR DNA fragment from W. Murcott and one clone from Tango were sequenced and compared with one another as well as with the corresponding regions of the left and right LTR of the CIRE1 as published by Rico-Cabanas and Martínez-Izquierdo (*Figure 3.5*). Comparisons showed an insertion of a single thymine base in the beginning of the W. Murcott and Tango clone sequences, while the left and right LTR regions of the published CIRE1 sequence lacked this insertion. An insertion of 8 base pairs was found in the 423.3Fx113 W. Murcott clone; this insertion was absent in the other clones as well as the published CIRE1 LTR sequences. The deletion of a thymine was found in the W. Murcott and Tango clones. Single nucleotide polymorphisms (SNPs) were found in all of the clones when compared to one another and when compared to the published CIRE1 LTR sequences (*Figure 3.5*); there was one SNP found in the published left-LTR region of CIRE1 that differed from all of the other compared sequences.

Genomic DNA Blots Used to Support qPCR Findings

Though the majority of DNA sequence found in each clone and in CIRE1 is identical, there is a small amount of variability (~5%) between the compared DNA sequences, due to point mutations found in each clone. This observed variability amongst the right-LTR retrotransposon sequences in CIRE1 will not affect qPCR amplification unless there is variability within the region of DNA targeted either by the forward or reverse primers (or

both) qPCR is a highly specific amplification reaction; if there is even slight base-pairing mismatch, the primer/s will not bind properly and the region of interest will not be amplified. Due to this possibility, genomic DNA blot analysis was performed to complement the qPCR analysis as it is less sensitive to variability in DNA sequence. Southern blots were also used to analyze other, less conserved regions of the CIRE1 LTR retrotransposon (integrase and reverse transcriptase genes) that could not be amplified well using qPCR, perhaps due to their high sequence variability.

Probes Selected for Hybridization to DNA Blots

For ^{32}P-labeled probe synthesis for DNA blot analysis, 388 to 471 bp DNA fragments were amplified from the right-LTR region of CIRE1 (*Figure 3.6A*) as well as the *integrase* gene (*Figure 3.6B*) and *reverse transcriptase* gene (*Figure 3.6C*). Three PCR amplification reactions were performed using W. Murcott as the DNA template for each reaction and a primer set targeting one of the three regions of interest. The DNA concentration of each PCR product was measured on the Nanodrop spectrophotometer; 500 ng of each was loaded into an agarose gel to ensure that the correct sized DNA fragment was amplified (*Figure 3.7A*). The control probe was amplified from the phospholipase D gene; this sequence was amplified using the same conditions and PCR program used with the retrotransposon primer sets only substituting primer sets targeting the phospholipase D. After measuring its DNA concentration with the Nanodrop spectrophotometer, 200 ng was loaded into an agarose gel to ensure that the correct sized DNA fragment was amplified (*Figure 3.7B*).

Southern Blot Analysis

Citrus varieties selected for Southern blot analysis were W. Murcott, Tango, and Abhayapuri Lime. Genomic DNA was digested with *EcoRI* or *HindIII* and fractionated on an agarose gel. Undigested genomic DNA served as a control (*Figure 3.8*). The remainder of digested genomic DNA (~3-5 µg per sample) was used to prepare Southern blots (*Figure 3.9*). A much fainter signal was detected from HindIII-digested W. Murcott as it had the least amount (~3 µg) of digested genomic DNA loaded. Also difficult to see are the positive/negative controls (LTR, INT, and RT PCR fragments) which are loaded at concentrations of 50-60 ng. DNA blots were hybridized to ^{32}P-labeled LTR, INT, RT, and *Phospholipase D* probes. Analysis usually produced a single strong signal for each *Citrus* variety studied (*Figures 3.10 A-D, 3.11 A,B*) as well as a weaker signal for its corresponding positive control PCR fragment (*Figures 3.10 B-D, 3.11 A,B*). When incubated for a longer period of time, some multiple, weaker signals could be seen using the LTR probe against the digested genomic DNA samples (*Figure 3.10C*) but the larger DNA fragment obscures the weaker signals from the smaller DNA fragments, making it difficult to determine with absolute certainty the number of DNA fragments hybridizing with the LTR probe in each digested genomic DNA sample. Weaker signals could not be seen in the *HindIII*-digested W. Murcott genomic DNA sample, which could be due to the fact that less genomic DNA (~3 µg) was loaded for this particular sample than were the other samples, which were loaded with about 5 µg per lane.

The weak multiple banding displayed by the *EcoRI*-digested W. Murcott genomic DNA sample suggests that these LTR probe-binding DNA fragments are indeed present in the genome, despite not being displayed in the *HindIII*-digested DNA sample of W. Murcott. Though it is difficult to determine with absolute certainty the number of signals generated by the LTR probe bound to smaller DNA fragments, the LTR probe signal (*Figure 3.10B*) is much stronger than that of the control probe (*Figure 3.10A*), INT probe (*Figure 3.10D*), or RT probe (*Figure 3.11B*) when exposed 18-24 hours (or 6 days for the RT probe). The greater signal for LTR suggests that this sequence is present in greater amounts within the genomes of each of the *Citrus* varieties than are the other ^{32}P-labeled DNA sequences. This comes as no great surprise since the LTR region in retrotransposons is known to be a very conserved DNA sequence, and thus would have very little variability in the sequence targeted by the probe. The INT and RT sequences, however, are not as conserved. The INT and RT sequences targeted by the probe could be more variable, meaning that the probe may not bind to sequences that have significantly altered from the original sequence. They both appear to be far less abundant in the *Citrus* genomes of the varieties studied when observing the strength of signal alone, with signal given off by the INT probe (*Figure 3.11A*) appearing to be about the same level as that given off by the RT probe (*Figure 3.11B*) This suggests that the INT DNA sequence is present in about the same amount as the RT DNA sequence, though not as great an amount as the LTR DNA sequence.

BLAST Alignment Against Reference Genomes

The Phytozome BLAST database containing draft assemblies of the *C. sinensis* and *C. clementina* genomes was used to determine how many CIRE1 LTR retrotransposons were present in these genomes. Four analyses were performed: (1) the full CIRE1 LTR retrotransposon sequence, (2) the full right-LTR sequence, (3) the full INT gene sequence, and (4) the full RT gene sequence (*Table 3.1*). The number of BLAST hits that were identified from the CIRE1 LTR retrotransposon was low (134 and 589). The right-LTR sequence, which is the most conserved of the three targeted CIRE1 components, generated the fewest hits of all of the sequences analyzed. Also, the amount of perfect sequence match (as indicated by the total number of "E-value 0" scaffold hits) against both *Citrus* reference genomes was low (less than 71) for all of the query sequences that were used. The same query sequences were used to BLAST against the Tango/W.Murcott assembly developed by Rebecca Sun (*Table 3.2*). An e-value threshold of 0.0001 was used in this analysis, the number of significant hits that were returned using this e-value ranged from 194-767 hits (*Table 3.2*). The number of perfect matches as indicated by an e-value score of 0 ranged from 6 to 25 for this comparison, and no perfect matches were detected for the LTR region.

Alignments of significant BLAST hits against the query sequence (*Figure 3.12A-D*) were generated for each sequence query: CIRE1 (*Figure 3.12A*), LTR (*Figure 3.12B*), INT (*Figure 3.12C*) and RT (*Figure 3.12D*). The most significant matches are shown in red, while the least significant matches are shown in green. Gray lines indicate gapped

regions between the query sequence and its hit. From this data, it appears that the full-length CIRE1 retrotransposon sequence has many significant matches (*Figure 3.12A*), however, nearly all of these cover only a fraction of the full-length sequence. Both the INT and RT full-length sequence queries (*Figure 3.12C, D)* had several significant matches that also covered the length of the query sequence. These represent the 6 perfect BLAST matches against INT and RT (*Table 3.2*).

Discussion

Results from qPCR analysis and BLAST searches against three genome sequence assemblies indicate that the CIRE1 LTR retrotransposon does not exist in a high copy number within the *Citrus* varieties studied, nor does there appear to be a significant difference in copy number of CIRE1 between W. Murcott and Tango. In the qPCR analysis, 1-12 copies of the right-LTR region were detected in *Citrus*, while 0-71 significant BLAST hits resulted from queries of various regions of CIRE1 against *Citrus* reference genomes. Past studies have shown that qPCR reactions are highly reproducible (Zemanick *et al* 2010, Chen *et al* 2009) and, according to Biosistemika, qPCR is sensitive enough to detect a single copy of DNA (http://www.qpcrexperience.com). It is also able to effectively quantify DNA over several orders of magnitude. Because qPCR is a highly sensitive reaction, it is imperative that the DNA sample does not contain any impurities that may inhibit the qPCR reaction. For this experiment, a Qiagen DNeasy Plant Mini Kit was used for genomic DNA extraction and purification. According to the Qiagen website, the kit uses a silica-gel–membrane along with simple spin procedures in order to isolate plant genomic DNA that is highly pure (http://www.qiagen.com). In addition, when quantifying the amount of DNA extracted, the Nanodrop Spectrophotometer has the capability to assess the purity of the DNA sample using the ratio of absorbance at 260nm over the absorbance at 280nm. Only DNA that met the satisfactory ratio requirement was used in the qPCR analysis.

The assembly of the *C. clementina* genome includes 296 Mb spread over 1,128 scaffolds and is based on whole genome shotgun approach with Sanger sequencing technology. The *C. sinensis* assembly includes 319 Mb in 12,574 scaffolds from predominantly 454 sequencing. The haploid *C. clementina* genome assembly is a high-quality 6.5X coverage draft with half of the genome accounted for by 27 scaffolds that are 3.3 Mb or longer. For the diploid, heterozygous *C. sinensis* genome, half of the assembly is accounted for by 236 scaffolds 251 kb or longer. Because the results of the qPCR analysis and the results of the BLAST against reference genomes analysis both indicate that CIRE1 is in a low copy number within the *Citrus* genomes analyzed, the conclusion that the CIRE1 retrotransposon exists in a lower copy number than previously stated by Rico-Cabanas and Martinez-Izquierdo (2006) is likely accurate due to the reliability of qPCR and the support given by the various genome assemblies.

There are several hypotheses that may explain the low copy number of the CIRE1 LTR retrotransposon. For example, Hickey proposed that sexual reproduction allows deleterious transposable elements to proliferate in populations, and the loss of sex prevents their spread and therefore would eventually result in a population free of such elements (Hickey *et al* 1982). Indeed, in citrus, as with most other cultivated fruit trees, a horticultural variety traces back to a single parent tree or individual mutant branch which has been multiplied by vegetative or asexual means, such as cuttage or graftage. Citrus varieties or cultivars are therefore clonal varieties. According to studies conducted by Arkhipova and Meselson (2000) the pattern of occurrence of LINE-like and gypsy like retrotransposons in sexual versus ancient asexual taxa supports Hickey's hypothesis. It is

therefore a possibility that the low frequency or complete absence of sexual reproduction in the varieties of *Citrus* studied has contributed to the low copy number of CIRE1 LTR retrotransposon found in each of them. However, sweet oranges have been asexually propagated for thousands of years (Liu and Deng 2007), and according to letters from Dr. El-Bachir Nadori, W. Murcott is derived from a seed from a Murcott tree, so he believed W. Murcott to be a chance hybrid of Murcott and a pollen parent whose identity remains unknown (http://www.citrusvariety.ucr.edu/citrus/wmurcott.html). W. Murcott was first imported to California from Morocco in 1985 and Clementine probably originated as a sexual seedling within the past 500 years (Hodgson 1967). It appears (*Tables 3.1, 3.2*) that CIRE1 copy number is higher in sweet orange than in Clementine, an observation which is not consistent with this prediction.

Another plausible idea, proposed by Vitte et al (2005) is that while plant genomes do at times undergo genome size increases through bursts of retrotransposition, there exists a counteracting process that eliminates the transposed copies from the genomes (Vitte *et al* 2004). This was clearly evident in DNA blots with barley using probes from the *RT, Int,* and LTR regions of the *BARE-1* retrotransposon. These studies showed that the ratio of LTR to internal regions varied greatly among species both within the genus *Hordeum* (Vicient *et al* 1999) and within wild barley (*H. spontaneum*, Kalendar *et al* 2000). In both cases, there was a negative correlation with the contribution of the *BARE-1* element to genome size (Vicient *et al*., 1999; Kalendar *et al*., 2000). These results suggested that, at least in barley and related species, a force that removes LTRs is active (Vitte *et al* 2005). Deletions have also been shown to be frequent for LTR retroelements

within the wheat genome (Wicker *et al* 2001, 2003), which suggested that formation of deletions by a mechanism independent of homologous recombination is key for DNA elimination in flowering plants (Vitte *et al* 2005).

Recently, DNA double-strand break repair was proposed to influence genome size and organization through homologous or illegitimate recombination. Kirik *et al* (2000) induced double-strand breaks in both *A. thaliana* and tobacco (two dicotyledonous plants species differing 120-fold in genome size) using the rare-cutting I-SceI restriction endonuclease and the sizes of the subsequent deletions were compared. Significant differences in deletion size were observed (Kirik *et al* 2000). First, the average deletion size was larger in tobacco than in Arabidopsis (1,341 bp compared to 920 bp). Second, insertions of filler sequences were associated with 40% of linked-ends cases in tobacco, whereas such insertions were not observed in Arabidopsis (Kirik *et al* 2000). Interestingly, such insertions of filler sequences have also been reported to be associated with deletions in maize, another plant with a large genome (Wessler et al., 1990). Given time, it is theoretically possible that LTR retrotransposons could be eliminated from the genomes of the *Citrus* varieties studied through the process of DNA repair and/or through mechanisms by which newly generated and inserted LTR retrotransposons are excised in order to conserve genome size. However, if the *Citrus* genome were to truly contain thousands of copies of CIRE1, the evolutionary process by which most of these copies would be eliminated from the genome would require hundreds of years. Since the study conducted by Rico-Cabanas and Martinez-Izquierdo was recent (2006) not enough time has passed for a large number of CIRE1 copies to be eliminated from the *Citrus* genome.

Rather, it is more likely that the particular Valencia cultivar studied by Rico-Cabanas and Martinez-Izquierdo underwent periods in which it was exposed to stressful conditions, which may have activated the CIRE1 retrotransposon and resulted in hundreds of additional copies. Such an event, while unlikely, cannot be ruled out as a possibility.

The method chosen to quantify the CIRE1 retrotransposon in the 2005 study conducted by Rico-Cabanas and Martinez-Izquierdo was slot blot hybridization, where genomic DNAs were immobilized on membranes, and hybridized to radiolabeled probes. In their study, Rico-Cabanas and Martinez-Izquierdo used *Citrus sinensis* genomic DNA (ranging from 50 to 500 ng) (Rico-Cabanas and Martinez-Izquierdo 2005). In addition, a series of dilutions containing PCR-amplified CIRE1 fragments (ranging from 0.1 to 2.5 ng) were also applied to the membrane (Rico-Cabanas and Martinez-Izquierdo 2005). ^{32}P-radiolabeled probes were made from the same PCR-amplified CIRE1 fragments (Rico-Cabanas and Martinez-Izquierdo 2005). Genomic copy number was calculated based on the hybridization signal from genomic relative to the CIRE1 reconstruction samples (Rico-Cabanas and Martinez-Izquierdo 2005). According to Kafatos et al (1979), this method can make semi-quantitative estimates of nucleic acid concentration, but it is not reliable as far as calculating exact nucleic acid concentrations in a mixture (Kafatos *et al* 1979). Commentary by Terry Brown regarding the dot blot hybridization method cautions that if impurities exist within the DNA, this can lead to unpredictable effects on hybridization, possibly reducing signal by blocking access to the hybridization sites, or increasing signal by trapping the probe (Brown 2001). Copy number reconstruction by dot blot analysis is therefore suspect if estimates of absolute concentrations of nucleic

acids in a mixture are based solely on the signal intensity compared with a control dilution series (Brown 2001). This comparison can only be reliable if both types of DNA are fully purified (Brown 2001). It is likely that the method chosen by Rico-Cabanas and Martinez-Izquierdo to determine the copy number of the CIRE1 retrotransposon within the *C. sinensis* genome was not a suitable one; slot blot hybridization should not be used when calculating precise copy numbers of specific DNA sequences as this method is not reliable beyond making a semi-quantitative estimate. The estimate of 2,200 full-length copies of the CIRE1 retrotransposon proposed by Rico-Cabanas and Martinez-Izquierdo to be present in the *C. sinensis* genome is not supported by the current study of CIRE1 copy number; it is possible that Rico-Cabanas and Martinez-Izquierdo erroneously overestimated the copy number of CIRE1 within the *C. sinensis* genome.

The suggestion that CIRE1 is an active LTR retrotransposon may actually increase the likelihood of purifying selection against newly introduced retrotransposons. According to findings by Navarro-Quezada and Schoen (2001), evidence of purifying selection was most pronounced in plant species with low estimated copy numbers of Ty1-copia elements. This finding is consistent with the hypothesis that high element turnover rates favor elements capable of transposition (Navarro-Quezada and Schoen 2001). In their study of RT sequences within Ty1-*copia* elements, simulations of RT sequence evolution indicate that it is incorrect to assume that low copy numbers of transposable elements are the product of reduced levels of element activity (Navarro-Quezada and Schoen 2001). They discovered that when there is strong selection against individuals harboring many elements, and or high per-element loss rates leading to low copy

numbers of elements, the vast majority of elements that comprise the equilibrium population of retrotransposons contain few nonsynonymous mutations and are transpositionally competent (Navarro-Quezada and Schoen 2001). While it has sometimes been argued that copy numbers of transposable elements can be interpreted as a direct reflection of activity levels, such as elements that are present in high copy numbers are equated with elements that have high transposition rates, the findings by Navarro-Quezada and Schoen suggest that this view is incorrect. Both the pattern of stronger purifying selection of RT sequences in genomes with fewer elements, and the findings from simulated RT sequence evolution suggest that there can be strong selection for conservation of function in elements that are present in relatively few copies per genome. That is to say, low copy numbers may reflect an underlying dynamic equilibirum, such that only the most transpositionally active elements are capable of long-term persistence. As a transcriptionally active retroelement, CIRE1 may exist in low copy numbers within the *Citrus* varieties studied simply because purifying selection is more pronounced in the presence of a sequentially intact, active retroelement.

The three analyses performed to estimate the copy number of CIRE1 LTR retrotransposon within the *Citrus* varieties studied all indicate that it exists in a low copy number (1-71). Results also suggest that there is no difference in CIRE1 copy number between the W. Murcott and Tango mandarin varieties, meaning that the retrotransposon was likely not activated through stress conditions brought on by gamma irradiation, which otherwise would have resulted in greater numbers of copies of CIRE1 within the Tango genome when compared to W. Murcott. Though CIRE1 may not have been

activated, this does not rule out the possibility of activation of other, as yet unidentified retrotransposons. Further experimentation as to whether or not other retrotransposons were activated and generated copies of themselves within the Tango genome in response to gamma radiation stress would be required to determine if this is indeed the case.

References

Altschul S.F., Madden T.L., Schäffer A.A., Zhang J., Zhang Z., Miller W., and Lipman D.J., 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. 25:3389-3402.

Arkhipova I. and Meselson M., 2000. Transposable elements in sexual and ancient asexual taxa. Proc. Natl. Acad. Sci. USA 97: 14473-14477.

Barkley N.A., Roose M.L., Krueger R.R., and Federici C.T., 2006. Assessing genetic diversity and population structure in a citrus germplasm collection utilizing simple sequence repeat markers (SSRs). Theor. Appl. Genet. 112: 1519-1531.

Biswas M.K., Baig M.N.R., Cheng Y-J., and Deng X-X., 2010. Retro-transposon based genetic similarity within the genus *Citrus* and its relatives. Genet. Resour. Crop Evol. 57: 963-972

Brown T., 2001. Dot and slot blotting of DNA. Current Protocols in Molecular Biology. 2.9.15–2.9.20

Campos E.T., Espinosa M.A.G., Warburton M.L., Varela A.S., Monter A.V., 2005. Characterization of mandarin (*Citrus* SPP.) using morphological and AFLP markers. Interciencia 30: 687-693.

Chen Y., Gelfond J.A.L., McManus L.M., and Shireman, P.K., 2009. Reproducibility of quantitative RT-PCR array in miRNA expression profiling and comparison with microarray analysis. BMC Genomics 10: 407.

De Felice B., Wilson R.R., Argenziano C., Kafantaris I., and Conicella C., 2008. A transcriptionally active *Copia*-like retroelement in *Citrus limon*. Cell. Molec. Biol. Lett. 14: 289-304.

Fang D., Krueger R.R., and Roose M.L., 1998. Phylogenetic relationships among selected citrus germplasm accessions revealed by Inter-simple sequence repeat (ISSR) markers. J. Amer. Soc. Hort. Sci. 123: 612-617.

Federici C.T., Roose M.L., and Scora R.W., 2000. RFLP analysis of the origin of *Citrus bergamia*, *Citrus jambhiri*, and *Citrus limonia*. Acta Hort. 535: 55-64.

Feschotte C., Jiang N., and Wessler S.R., 2002. Plant transposable elements: where genetics meets genomics. Nature Rev. Genetics 3: 320-341.

Froelicher Y., Dambier D., Bassene J.B., Costantino G., Lotfy S., Didout C., Beaumont V., Brottier P., Risterucci A.M., Luro F., and Ollitrault P., 2008. Characterization of microsatellite markers in mandarin orange (*Citrus reticulata* Blanco). Molec. Ecol. Resources 8: 119–122.

Grandbastien M.A., Lucas H., Morel J.B., Mhiri C., Vernhettes S., and Casacuberta J.M., 1997. The expression of the tobacco *Tnt1* retrotransposon is linked to plant defense responses. Genetica 100: 241–252.

Grandbastien M.A., Audeon C., Bonnivard E., Casacuberta J.M., Chalhoub B., Costa A.-P.P., Le Q.H., Melayah D., Petit M., Poncet C., Tam S.M., Van Sluys M.-A., and Mhiria C., 2005. Stress activation and genomic impact of *Tnt1* retrotransposons in Solanaceae. Cytogenet. Genome Res. 110: 229–241.

Gupta P.K. and Rustgi S., 2004. Molecular markers from the transcribed/expressed region of the genome in higher plants. Funct. Integr. Genomics 4: 139–162.

Hickey D.A., 1982. Selfish DNA: a sexually-transmitted nuclear parasite. Genetics 101: 519-531.

Hodgson R.W., 1967. Horticultural varieties of citrus, In: Reuther, W., Webber, H.J., and Batchelor, L.D. (eds.). The Citrus Industry, vol. 1. Univ. Calif., Division of Agricultural Sciences. Berkeley.

Kafatos F.C., Jones C.W., and Efstratiadist A., 1979. Detemination of nucleic acid sequence homologies and relative concentrations by a dot hybridization procedure. Nucleic Acids Res. 7: 1541-1552.

Kalendar R., Tanskanen J., Immonen S., Nevo E., and Schulman A.H., 2000. Genome evolution of wild barley (*Hordeum spontaneum*) by *BARE-1* retrotransposon dynamics in response to sharp microclimatic divergence. Proc. Natl. Acad. Sci. USA 97: 6603–6607.

Kalendar R. and Schulman A.H., 2007. IRAP and REMAP for retrotransposon-based genotyping and fingerprinting. Nature Protocols 1: 2478-2484.

Kalendar R., Flavell A.J., Ellis T.H.N., Sjakste T., Moisy C., and Schulman A.H., 2011. Analysis of plant diversity with retrotransposon-based molecular markers. Heredity 106: 520–530.

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

Kumar A. and Bennetzen J.L., 1999. Plant retrotransposons. Ann. Rev. Genet. 33: 479-532.

Liu Y-Z. and Deng X-X., 2007. Citrus breeding and genetics in China. The Asian and Australian J. of Plant Sci. Biotech. 1: 23-28.

Manetti M.E., Rossi M., Costa A.P.P., Clausen A.M., and Van Sluys M.-A., 2007. Radiation of the Tnt1 retrotransposon superfamily in three Solanaceae genera. BMC Evolutionary Biology 7: 1-12.

Navarro-Quezada A. and Schoen D.J., 2001. Sequence evolution and copy number of Ty1-*copia* retrotransposons in diverse plant genomes. Proc. Natl. Acad. Sci. USA 90 : 268-273.

Nicolosi E., Deng Z.N., Gentile A., La Malfa S., Continella G., and Tribulato E., 2000. Citrus phylogeny and genetic origin of important species as investigated by molecular markers. Theor. Appl. Genet. 100: 1155-1166.

Pereira V., 2004. Insertion bias and purifying selection of retrotransposons in the *Arabidopsis thaliana* genome. Genome Biol. 5: R79.

Rico-Cabanas L. and Martínez-Izquierdo J.A., 2006. CIRE1, a novel transcriptionally active *Ty1-copia* retrotransposon from *Citrus sinensis*. Mol. Genet. Genomics 277: 365–377.

Roose M.L. and Williams T.E., 2007. Mandarin tree named 'Tango'. US Plant Patent PP17,863.

Sambrook J., Fritsch E.M., and Maniatis T., 1989. Molecular cloning: a laboratory manual, 2nd edition, Cold Spring Harbor Laboratory, Cold Spring Harbor Labor, New York.

Sugarawa K. and Oowada A., 1995. Identification of *Citrus* chimeras by RAPD analysis. HortSci. 30: 1276-1278.

Takeda S., Sugimoto K., Otsuki H., and Hirochika H., 1998. Transcriptional activation of the tobacco retrotransposon *Tto1* by wounding and methyl jasmonate. Plant Mol. Biol. 36: 365–376.

University of California, Riverside, citrus variety collection. http://www.citrusvariety.ucr.edu/citrus/wmurcott.html. Accessed November 4, 2011.

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

Vitte C. and Panaud O., 2005. LTR retrotransposons and flowering plant genome size: emergence of the increase/decrease model. Cytogenet. Genome Res. 110: 91–107.

Voytas D., and Ausubel F.M., 1988. A copia-like transposable element found in *Arabidopsis thaliana.* Nature 336: 242-244.

Wessler S.R., Tarpley A., Purugganan M., Spell M., and Okagaki R., 1990. Filler DNA is associated with spontaneous deletions in maize. Proc. Natl. Acad. Sci. USA 87:8731–8735.

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

Wicker T., Yahiaoui N., Guyot R., Schlagenhauf E., Liu Z.D., Dubcovsky J., Keller B., 2003. Rapid genome divergence at orthologous low molecular weight *Glutenin* loci of the A and Am genomes of wheat. Plant Cell 15: 1186–1197.

Zemanick E.T., Wagner B.D., Sagel S.D., Stevens M.J., Accurso F.J., and Harris J.K., 2010. Reliability of quantitative real-time pcr for bacterial detection in cystic fibrosis airway specimens. PLoS One 5 (11): (e15101).

Table 3.1 BLAST Against *C. Clementina*/*C. Sinensis* Reference Genomes From Phytozome Website

Query region	Target Genome	Query bp length	Number of scaffold hits with threshold E-value <10	Total # of scaffold hits with E-value of 0
CIRE1 LTR retrotransposon	C. clementina	5044	134	38
CIRE1 LTR retrotransposon	C. sinensis	5044	589	71
Right LTR region of CIRE1	C. clementina	461	19	8
Right LTR region of CIRE1	C. sinensis	461	24	4
CIRE1 INT gene	C. clementina	771	106	12
CIRE1 INT gene	C. sinensis	771	391	12
CIRE1 RT gene	C. clementina	729	122	12
CIRE1 RT gene	C. sinensis	729	430	9
Phospholipase D	C. clementina	779	5	1
Phospholipase D	C. sinensis	779	5	1

BLAST results from various regions of CIRE1 LTR retrotransposon against Citrus genomes sinensis (target length 319231331) and clementine (target length 295550349); an E-value threshold parameter value of -10 was selected in order to filter out less significant scaffold hits; most regions of CIRE1 received over 100 hits, with the right LTR region receiving the fewest number of hits

Table 3.2 BLAST Against W.Murcott/Tango Sequences Assembled Using Abyss

Query region	Query bp length	Number of BLAST hits with threshold E-value^(0.0001)	Total # of BLAST hits with E-value of 0
CIRE1 LTR retrotransposon	5044	767	25
Right LTR region of CIRE1	461	194	0
CIRE1 INT gene	771	271	6
CIRE1 RT gene	729	282	6
Phospholipase D	779	15	3

BLAST results from various regions of CIRE1 LTR retrotransposon against Tango/W.Murcott sequence assembly; an E-value threshold parameter value of 0.0001 was selected in order to filter out less significant sequence hits; most regions of CIRE1 received over 100 hits, with the right LTR region receiving the fewest number of hits; perfect matches indicated by an e-value score of 0 are shown in the far right column

AAACAGAGGCTGACGAATGGTGCCTCATACTGTGCTAATATCGTCTTCATTTGTTC
ATCATTTTGCATAGAATCTGTTTCATTGTGATCTGTGCATAATCTTCTGTAAATCAC
CTGTGATTGGGTTAGTTCTTGATCAATAATAATGAGTAAATTTTCCTTCTAAGTTTG
CTGGTCTGTGTCTTTTGTTTTGTCTCTGGTTTAGCTTCTTTGTTCTGGTTTTTCTTGA
TTCGATTGTGTCCT TGG

Figure 3.1 Right-LTR Region Targted in QPCR Analysis: *Shown here is the right-LTR region targeted in the qPCR analysis of the CIRE1 LTR retrotransposon; highlighted in yellow are the forward and reverse primer sequences of primerset 2F/2R used in the qPCR reaction, flanking the DNA sequence amplified between both primers*

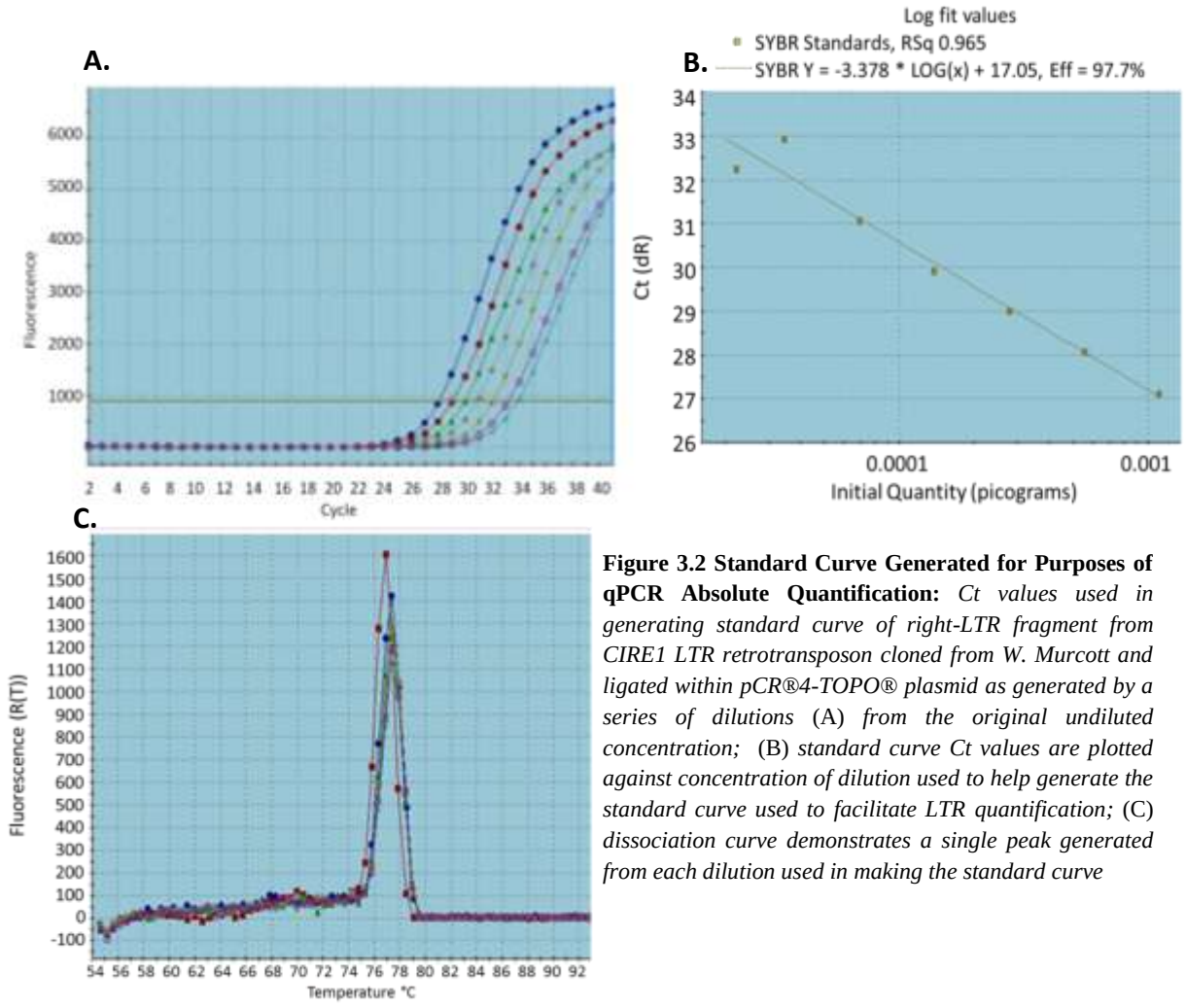


Figure 3.2 Standard Curve Generated for Purposes of qPCR Absolute Quantification: *Ct values used in generating standard curve of right-LTR fragment from CIRE1 LTR retrotransposon cloned from W. Murcott and ligated within pCR®4-TOPO® plasmid as generated by a series of dilutions* (A) *from the original undiluted concentration;* (B) *standard curve Ct values are plotted against concentration of dilution used to help generate the standard curve used to facilitate LTR quantification;* (C) *dissociation curve demonstrates a single peak generated from each dilution used in making the standard curve*

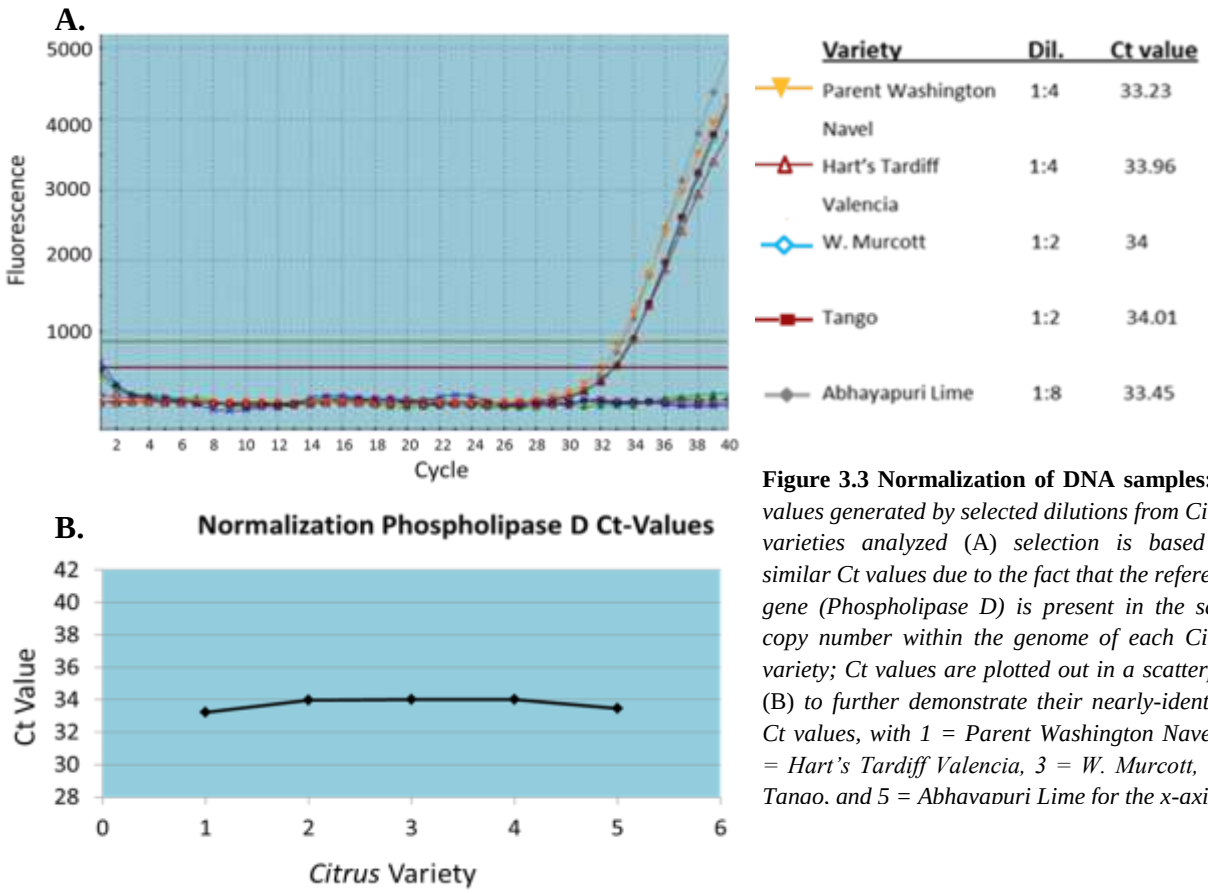


Figure 3.3 Normalization of DNA samples: *Ct values generated by selected dilutions from Citrus varieties analyzed* (A) *selection is based on similar Ct values due to the fact that the reference gene (Phospholipase D) is present in the same copy number within the genome of each Citrus variety; Ct values are plotted out in a scatterplot* (B) *to further demonstrate their nearly-identical Ct values, with 1 = Parent Washington Navel, 2 = Hart's Tardiff Valencia, 3 = W. Murcott, 4 = Tangao, and 5 = Abhayapuri Lime for the x-axis*

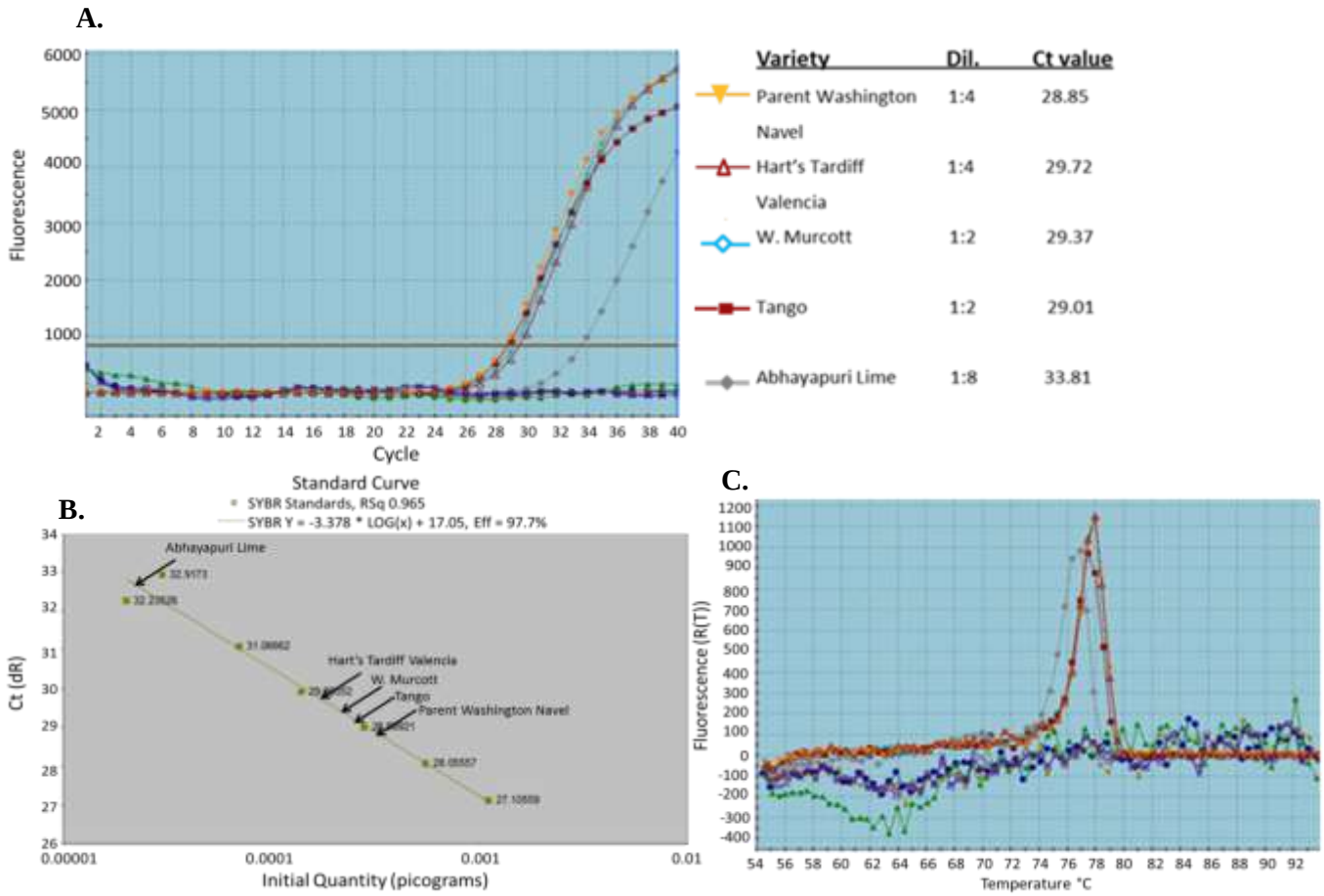


Figure 3.4 Generation of Right-LTR Ct Values From Five Citrus Varieties: (A) *qPCR analysis of right- LTR fragment in CIRE1 LTR retrotransposon; (B) Citrus varieties plotted on the standard curve based on their individual Ct values; (C) dissociation curves for products generated by each diluted sample where a single peak generated by each sample signifies that only a single sequence is present; the similar Tm to that observed in figure 3.2C suggests that the DNA products shown are similar to that shown in figure 3.2C.*

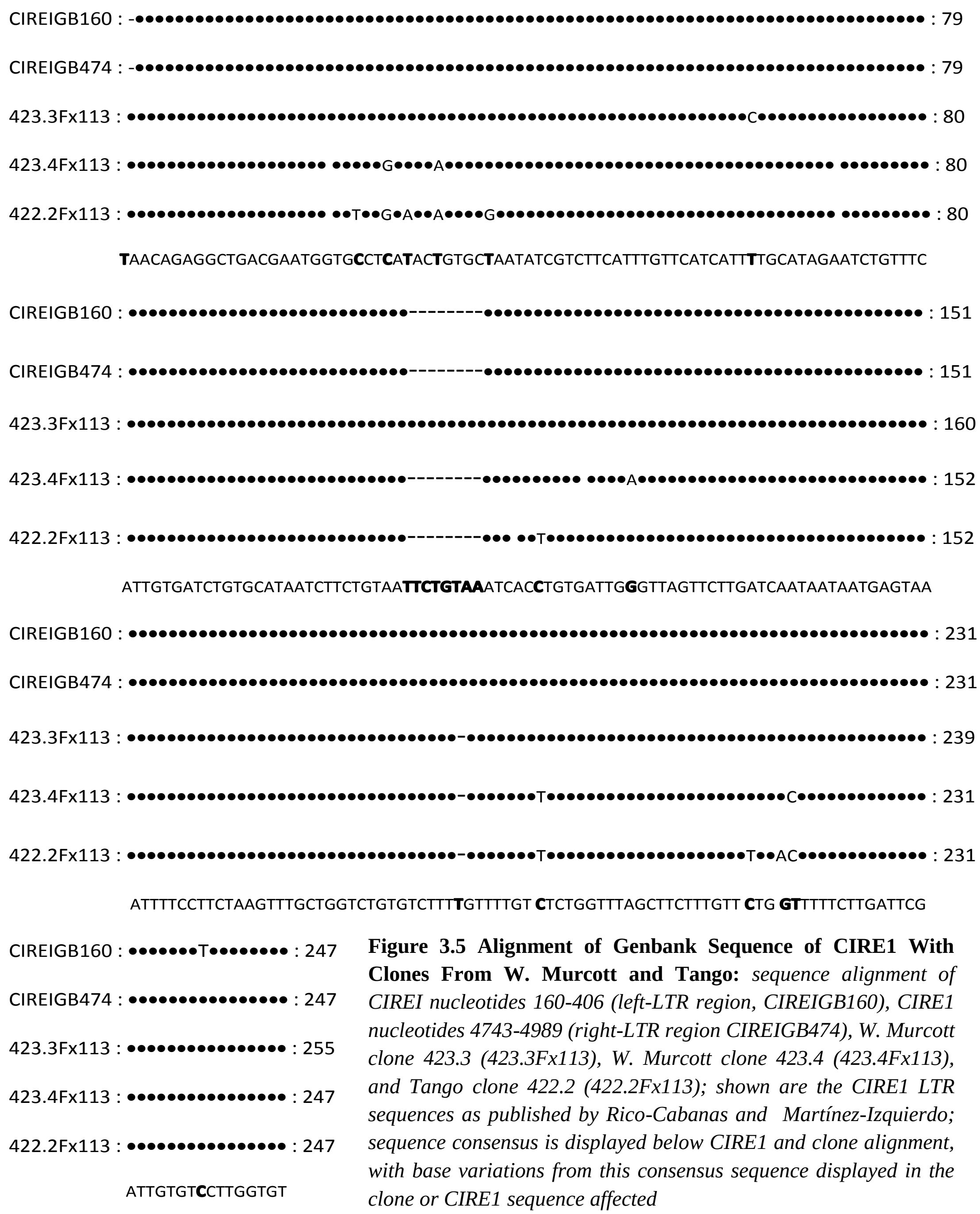


Figure 3.5 Alignment of Genbank Sequence of CIRE1 With Clones From W. Murcott and Tango: *sequence alignment of CIREI nucleotides 160-406 (left-LTR region, CIREIGB160), CIRE1 nucleotides 4743-4989 (right-LTR region CIREIGB474), W. Murcott clone 423.3 (423.3Fx113), W. Murcott clone 423.4 (423.4Fx113), and Tango clone 422.2 (422.2Fx113); shown are the CIRE1 LTR sequences as published by Rico-Cabanas and Martínez-Izquierdo; sequence consensus is displayed below CIRE1 and clone alignment, with base variations from this consensus sequence displayed in the clone or CIRE1 sequence affected*

A.

GAGCTGTTGTGCAGATTGACTTCAGTTGATGGGGTCTATTTGATATCTTCCGCTGACGTGGCATG
AATTCGTTATAATCTGTTATAACGAAGCCCAGCACACTGATATAAAACAGAGGCTGACGAATGG
TGCCTCATACTGTGCTAATATCGTCTTCATTTGTTCATCATTTTGCATAGAATCTGTTTCATTGTGA
TCTGTGCATAATCTTCTGTAAATCACCTGTGATTGGGTTAGTTCTTGATCAATAATAATGAGTAAA
TTTTCCTTCTAAGTTTGCTGGTCTGTGTCTTTTGTTTTGTCTCTGGTTTAGCTTCTTTGTTCTGGTTT
TTCTTGATTCGATTGTGTCCTTGGTGTTTCTACTCACCAAGATTCGTGAGTCTTGCTG

B.

GATAGAATCCAGCAACTGGAGTTTTGTGAAAACTGTATTTTTGGCAAATCACACAGGTCCAAGTT
TAACAAAGGGGAGCATATGTCTAAACAAGTTTTGGACTATGCACATATTGACCATTGGGGACCT
GCTCAGGTCCCTTCCTTGAGTGGTGGCAGATATTTCATGTCACTTATTGATGATTACTCAAGGAA
AGTATGGATCTACATACTCAAGATCAAAGATCAAGCTTTGGAAAAATTCAAGGTCTGGAAATCAT
TGGTAGAAAATCAGTCTGACTTCAAGCTGAAATGCTTGAGGACTGATAATGGCTTGGAATTTTGT
AGCAAAGTATTTGAAGAATATTGCCAAAAACATGGCATAAAGAGGCACAAAACAGTGAGGTTCA
CCCCCCAGCAGAATGGCTTGGCTGAGAGGATGAACAGGACCTTGGTGGATAAAACCAGATGTA
TGCTGATAAATTCAAAGCT

C.

GAAGGTGTAGATTTCAATGAAATTTTTTCTCCTGTAGTGAAGCACAGTTCAATAAGAATCTTATT
AGCCATGGTAGCTTTACTTGACTTGGAGCTGGAGCAGATGGATGTAAAAACAGCCTTCCTTCAT
GGTAATCTTGAAGAACAAATATTGATGGCACAGCCTGAAGGATTTGAGTGTAAGGATAAGGAG
GATTATGTGTGTCTGCTGCATAAGTCCTTGTATGGACTCAAGCAATCCCCTCGACAGTGGTACAG
AAGGTTTGATGATTTTATGGTCTCAAAGGGCTATCGCAGAAGCAGATATGACAGCTGTGTTTACT
TTGGTGGATCAGATCAAGGTGGAGTTGCTTATTTGCTTCTGTATGTTGACGATATGTTAATTGCA
AGCAAATACAAGTCAG

Figure 3.6 Targeted Regions of CIRE1: *CIRE1 LTR retrotransposon sequence regions targeted as probes to be used in Southern blot, including 388 bp right-LTR region* (A), *470 bp integrase gene region* (B), *and 403 bp reverse transcriptase gene region* (C); *sequences highlighted in yellow are primers, sequences highlighted in light blue is the DNA sequence amplified between primers*

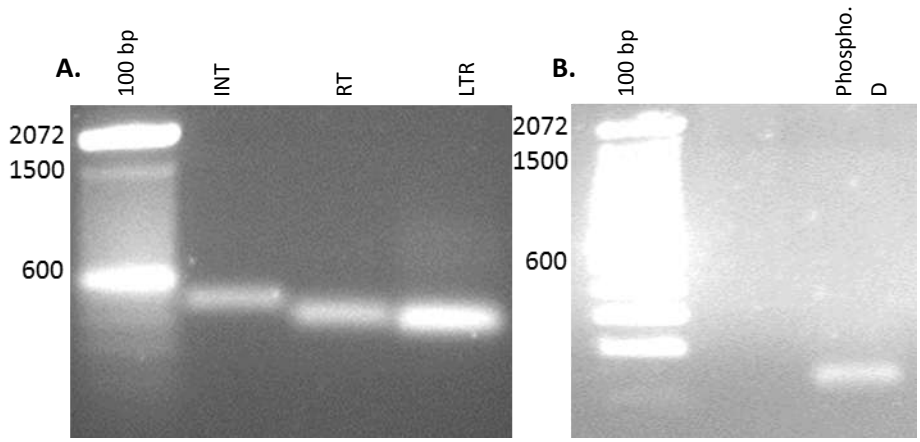


Figure 3.7 PCR Fragments Used as Probes in Southern Blot Anaysis: *0.9% agarose gel incorporated with EtBr was used to view PCR amplified products; shown* (A) *are 471-bp integrase, 403-bp reverse transcriptase, and 388-bp right-LTR DNA fragment; shown to the far left is 550 ng of loaded DNA 100-bp ladder; shown* (B) *is 150-bp Phospholipase D; shown to the far left is 550 ng of loaded DNA 100-bp ladder*

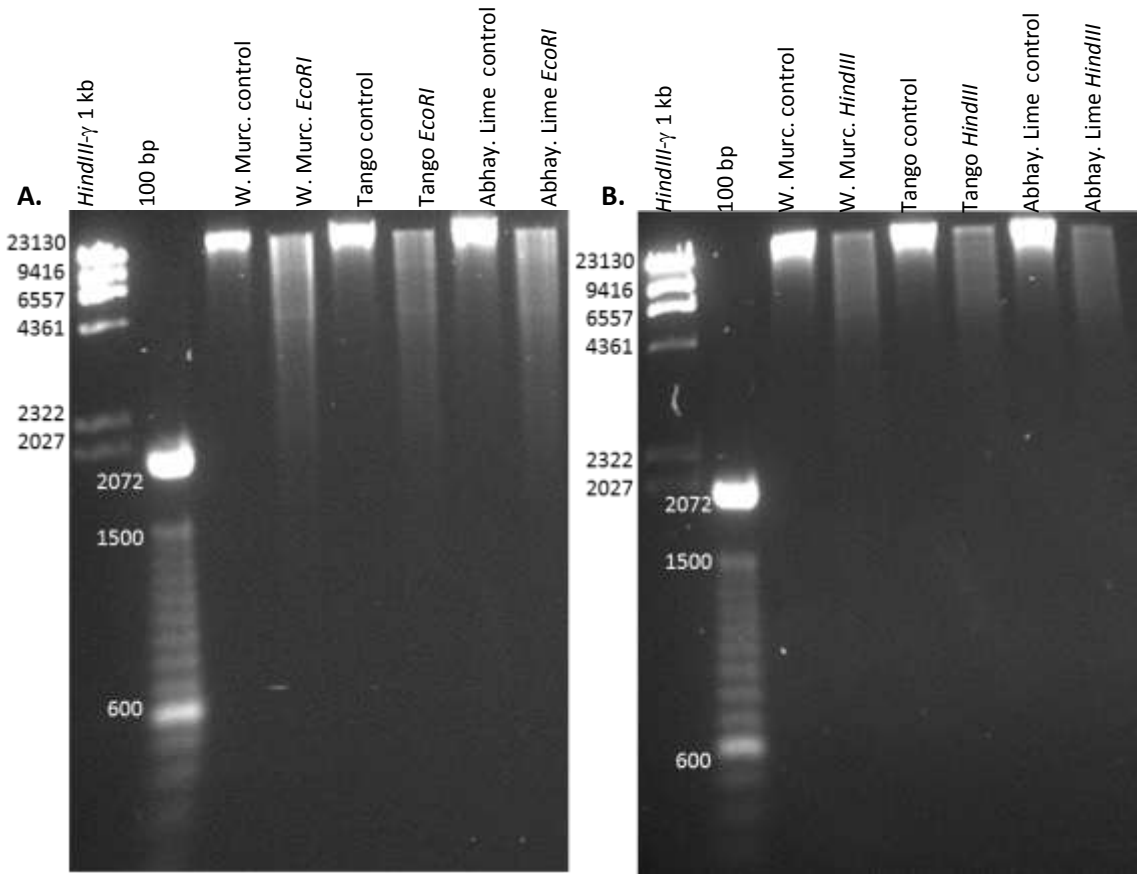


Figure 3.8: *Citrus genomic DNA digests compared with their undigested genomic controls; shown are EcoRI* (A) *and HindIII* (B) *digests along with HindIII-digested λ phage and 100 bp DNA ladders; about 150 ng of control and digested genomic DNA was loaded per lane; gels were run at 90V for about 40 min followed by 40V for about 2.2 hrs*

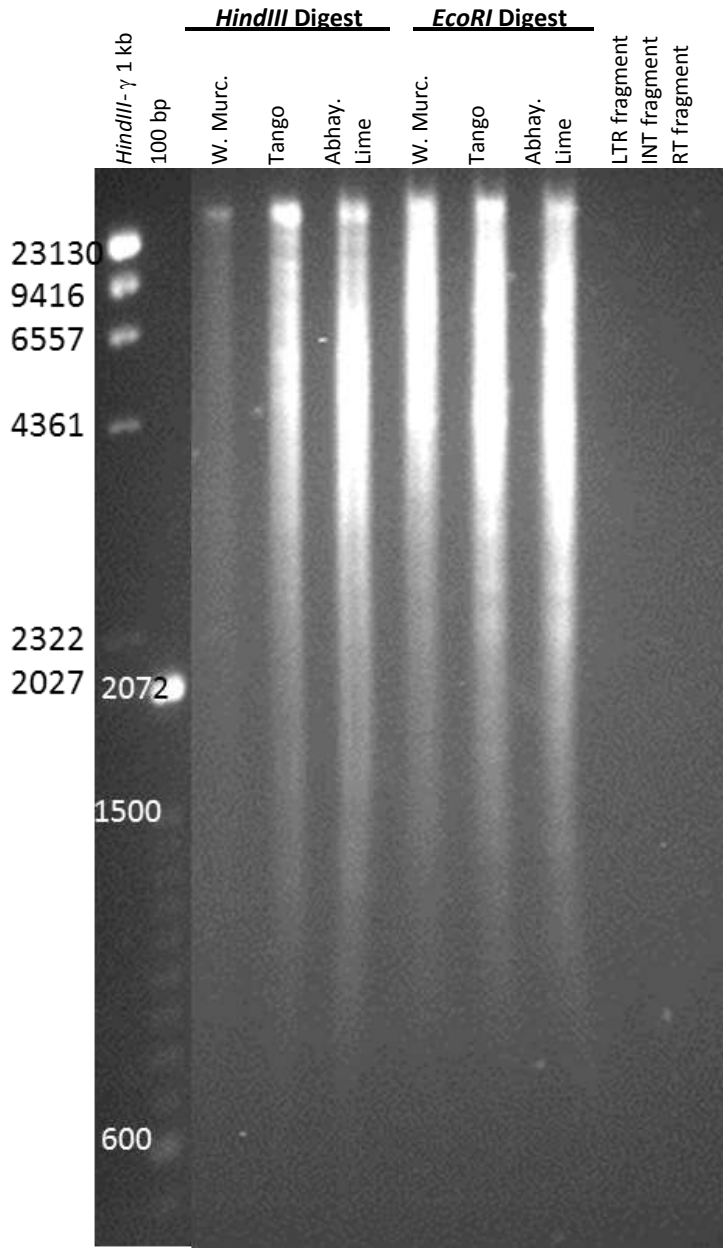


Figure 3.9: *HindIII and EcoRI digests of Citrus genomic DNA separated on 0.9% agarose gel containing EtBr, the digests and ladders were run at 100V for 2 hrs followed by 40V for 18hrs*

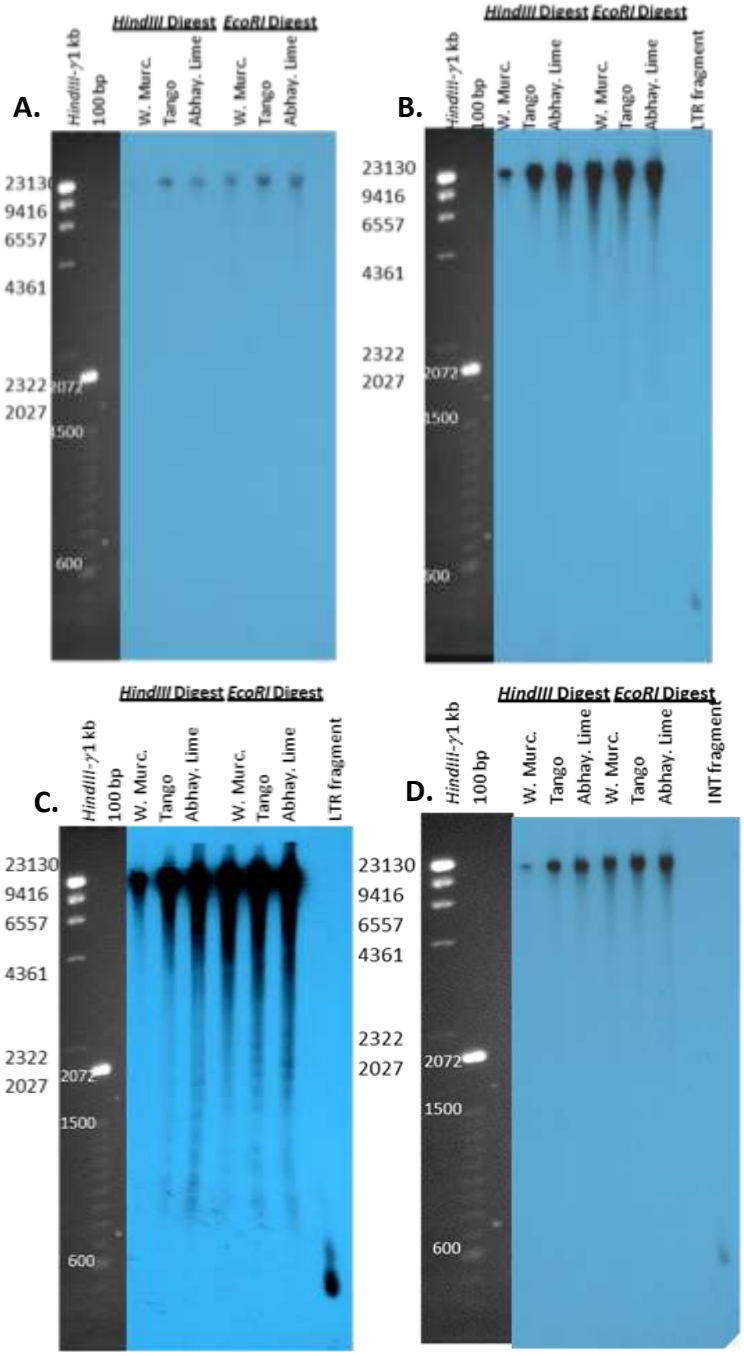


Figure 3.10 Southern Blots Using Probes *Phospholipase D*, LTR, and *INT*: *Shown are Southern blot film developments of* ^{32}*P-labeled phospholipase D probe incubated 19 hrs* (A) *,right-LTR CIRE1 probe incubated 18 hrs* (B), *and 3 days* (C), *and INT CIRE1 probe incubated 18 hrs* (D)

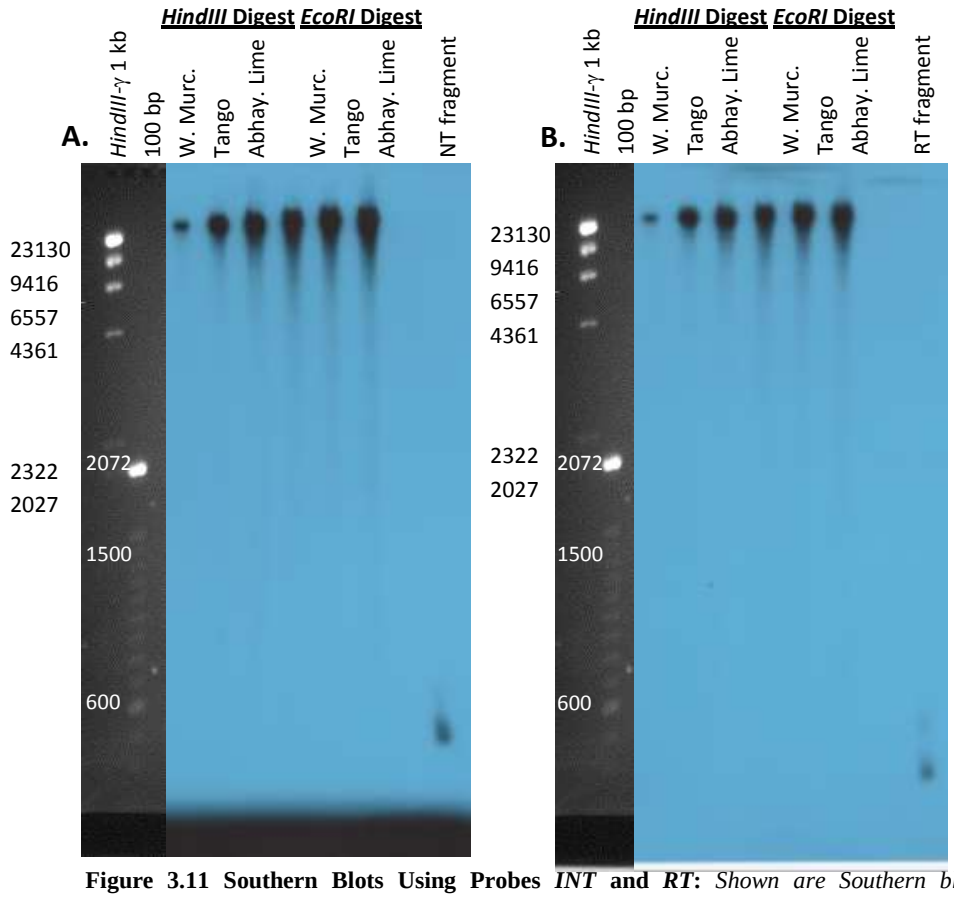


Figure 3.11 Southern Blots Using Probes *INT* and *RT*: *Shown are Southern blot film developments of* ^{32}*P-labeled INT CIRE1 probe incubated 3 days* (A) *and RT CIRE1 probe incubated 6 days*(B); *overall signal strength of the probes bound to genomic DNA fragments is similar*

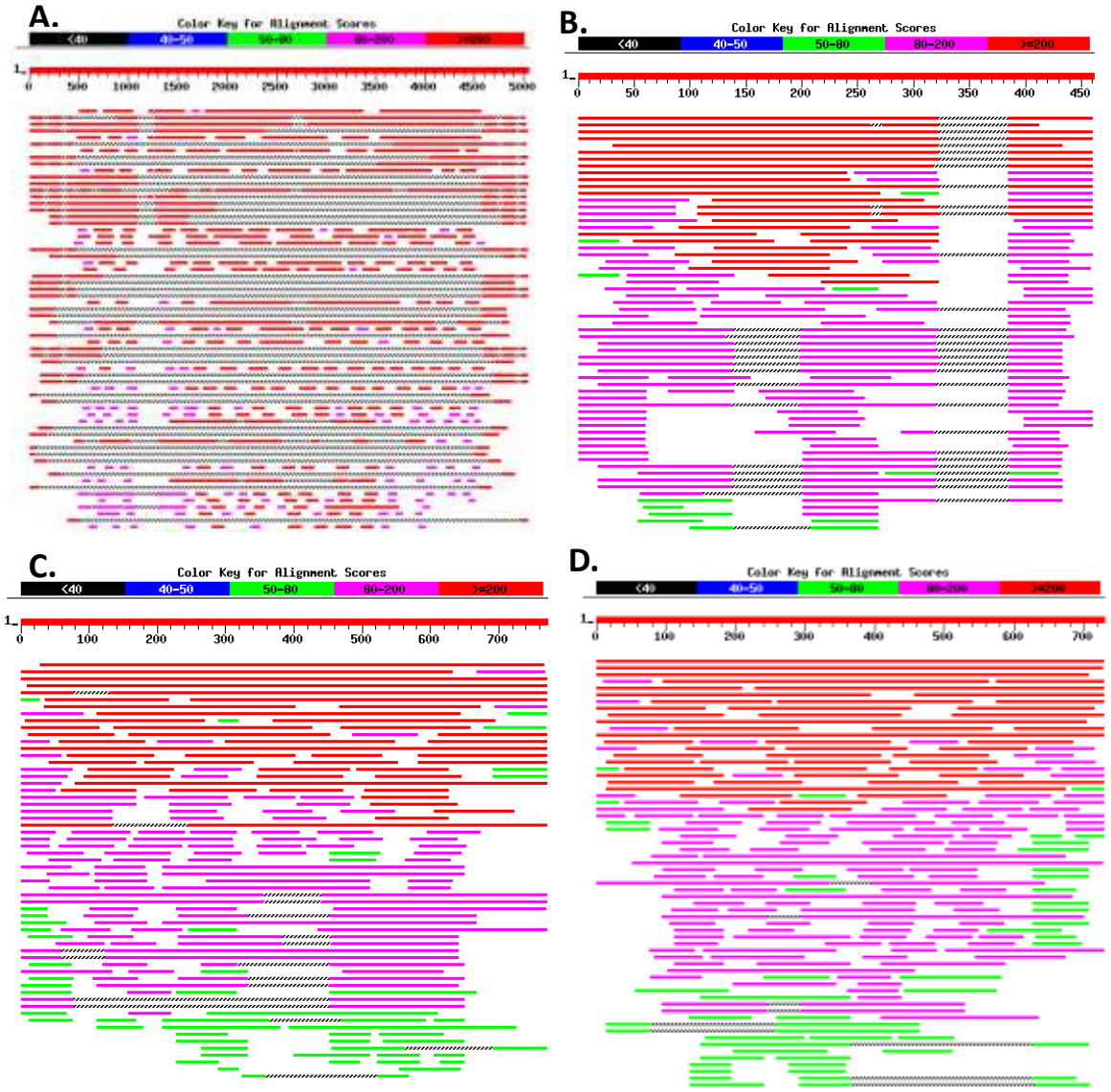


Figure 3.12 BLAST sequence alignments: *Sequence alignments of CIRE1 sequences to Tango/W.Murcott assembly. Strongest sequence matches to the query sequence are shown in red with progressively weaker matches in pink, green and black. Alignment s are shown for (A) full CIRE1 LTR retrotransposon, (B) full-length right LTR sequence, (C) full-length INT gene, and (D) full-length RT gene.*

Chapter 4

Analysis and comparison of genome-wide cytosine DNA methylation between W. Murcott and Tango mandarins using LUMA and methylation-sensitive AFLP techniques

Abstract

Conventional citrus breeding techniques that are used in order to produce high fruit quality traits, such as seedlessness, can be slow and difficult due to the complex biology of *Citrus.* Conventional breeding efforts are also limited by the little that is known about the mode of inheritance of desirable traits. However, citrus breeders have begun to take advantage of inducing spontaneous mutations in order to improve their crops, utilizing mutagenic tools such as radiation. Radiation treatment may induce genomic instability, which can be epigenetic in nature. Methylation of cytosines within a genomic DNA sequence is a common epigenetic mechanism, and it may repress gene expression if the methylation is located within the vicinity of genes. Conversely, previously silenced genes may become expressed if nearby methylated cytosines were to become unmethylated. DNA damaging agents such as radiation are believed to be involved in the alteration of DNA methylation patters, thus explaining their genome destabilizing effects. Budwood from a seedy W. Murcott mandarin was treated with gamma radiation in an effort to induce mutation. The result was a high quality, seedless Tango mandarin. A possible underlying cause of the seedless phenotype seen in Tango could be a change in the

methylation of its cytosines, which in turn could change the transcription pattern of its genes, which could result in a seedless phenotype. The purpose of this study was to investigate and compare cytosine methylation within the genomes of W. Murcott and Tango *Citrus* trees, the hypothesis being that an alteration in cytosine methylation should be found in Tango when compared to W. Murcott due to genomic instability induced by radiation treatment. Two methods were used in this analysis: LUMA pyrosequencing, which measured the percentage of methylated cytosines in each genome, and methylation-sensitive amplified fragment length polymorphism (MS-AFLP), which targeted methylated and unmethylated cytosine sites within each genome. Both analyses used methylation-sensitive (*HpaII*) and insensitive (*MspI*) enzymes to digest the DNA. DNA was extracted from gynoecium and young leaf tissue from W. Murcott and Tango trees for LUMA analysis, while DNA extracted from young leaf tissue from W. Murcott and Tango trees was used in the MS-AFLP analysis. Results from the LUMA analysis revealed that the percentage of methylated cytosines range from 53.1% in Tango gynoecium tissue to 58.6% in W. Murcott leaf tissue. Using the Student t-test, all comparisons of methylated DNA from the tissues studied between W. Murcott and Tango were found to have a P-value higher than 0.05, the cutoff value for significance. When averaged over genotypes, gynoecium DNA had significantly less methylated cytosine (53.3%) than young leaf DNA (57.2%). Results from the MS-AFLP analysis showed that all of the targeted methylated/unmethylated cytosine sites showed no difference in banding pattern between W. Murcott and Tango DNA. Of the total number of target sites analyzed within the *Citrus* genome (1457) 52% were estimated to be methylated, while

48% were estimated to be unmethylated. These results indicate that there is no significant difference in the percentage of methylated cytosines between W. Murcott and Tango genomes, nor is there a significant difference in the methylated/unmethylated cytosine patterns of sites that were targeted within W. Murcott and Tango genomes.

Introduction

The focus of Citrus improvement has been on high fruit qualities (e.g., seedless, easy peeling and mandarin types) and disease resistance (Sutarto *et al* 2009). These improvement programs have been implemented through standard breeding (hybridization), mutation breeding, and biotechnological techniques such as protoplast fusion to produce somatic hybrids (Sutarto *et al* 2009, de Assis Alves Mourao Filho *et al* 2008). Hybridization between species of the predominantly diploid *Citrus* genus often results in viable hybrids that can display valuable traits (Kamiri *et al* 2011). However, conventional citrus breeding techniques are often slow and difficult due to the multifaceted reproductive biology of *Citrus*, as well as complications such as high heterozygosity, apomixis, polyembryony and a long juvenile period (JinPing *et al* 2009). Moreover, conventional breeding efforts are restricted due to limited knowledge of the mode of inheritance of desirable traits (Vardi *et al* 2008). Several commercially valuable citrus cultivars, including Washington navel orange, Marsh grapefruit, Shamouti Orange, and Salustiana orange, came about through bud "sprout" mutations (Sutarto *et al* 2009). Tools for inducing mutations such as radiation or chemical mutagens are desirable for the expansion of variability in crop species because spontaneous mutations occur very infrequently (Sutarto *et al* 2009). An irradiation approach is useful as far as inducing complete, or very nearly complete, seedlessness in newly selected citrus cultivars (Vardi *et al* 2008). The after-effects of radiation treatment, which often contribute to genomic

instability to some degree, have been suggested to be epigenetic in nature (Nagar *et al* 2003).

As the ability to recognize and understand chromosome-based control systems has developed, this newfound knowledge has been exploited by plant breeders through the use of epigenetic sources of variation in producing new and improved hybrids (Hollick and Springer 2009). Epigenetics involves the change in gene expression through mechanisms other than change in genomic DNA sequence, such as DNA methylation or histone deacetylation, both of which usually repress gene expression without altering the sequence of the silenced genes.

Epigenetic regulatory systems were first studied extensively in maize; the discovery of these systems largely arose from the meticulous study of genetic features found in maize that demonstrate inheritance patterns that could not be sufficiently explained from Mendelian genetic principles (Hollick and Springer 2009). Dr. Barbara McClintock (1951) was the first to define a nuclear system in which its main function is to repress transposon activity (Hollick and Springer 2009). Her discoveries stemmed from experiments intended to stimulate segmental deficiencies using a chromosome type breakage-fusion-bridge (BFB) cycle (Hollick and Springer 2009). McClintock discovered that BFB cycles generated in early embryogenesis resulted in two frequent outcomes: the first being that the chromosome complement was often rearranged at positions not predicted to be encompassed in the BFB cycle, and the second that there was a substantial increase in the frequency of new, mutable phenotypes found among self-pollinated

progeny (McClintock 1951). McClintock's findings led to the conclusion that modifications to the heterochromatin caused the novel, unstable phenotypic variations to form in her cultures (McClintock 1951).

Silent transposable elements are methylated in plants (Rabinowicz *et al* 2003), and can be activated in methylation- deficient mutants (Rabinowicz *et al* 2003). Terminal inverted repeats (TIR)-type elements and their offshoots have the ability to regulate or impede standard means of gene regulation in a variety of ways (Hollick and Springer 2009). In such cases as a coding sequence insertion, the element either impedes mRNA maturation or it generates a changed reading frame that alters normal protein function (Hollick and Springer 2009). Numerous other cases exist in which a transposon is capable of gene expression regulation either directly, such as transposon-induced antisense RNA which can silence neighboring genes (Puig *et al* 2004), or indirectly via epigenetic alterations, such as epigenetic variation in the retrotransposon *Tos*17 activity in rice varieties, which in turn can affect the expression of nearby genes should an activated *Tos*17 insert itself within or in close proximity to the genes (Cheng *et al* 2006). Studies in which alterations, or epigenetic modifications, of a transposon family control the expression of genes located near autonomous or non-autonomous transposable elements impart the foundation for McClintock's portrayal of transposons as "controlling elements" (Hollick and Springer 2009). A majority of the well-researched transposons "cycle" between active and inactive conditions. These conditional changes are associated with changes in the levels of cytosine methylation (Hollick and Springer 2009). Normally, retrotransposons found in maize exist in heterochromatic regions in which the

DNA has high levels of cytosine methylation (Hollick and Springer 2009). In recent studies, plant mutants impaired in epigenetic repression indicate that deficiencies in methylation result in retrotransposon reactivation (Hollick and Springer 2009).

In plants and mammals, heterochromatic DNA is heavily methylated at cytosine residues (mC) and is linked with deacetylated histones H3 methylated at lysine 9 (H3K9me) (Tittel-Elmer *et al* 2010). DNA methylation levels are lower in euchromatin. Conversion from heterochromatin to euchromatin occurs when H3 acquires acetylation and methylation at lysine 4 (H3K4me) losing H3K9me during the process (Tittel-Elmer *et al* 2010). Knowledge of the magnitude and interaction of these epigenomic sources could add strongly to molecular genetic tools used in plant improvement (Hollick and Springer 2009). Methylation of cytosine residues characterizes the usual alteration of nuclear genomes in higher eukaryotes (Fojtová *et al* 2000). Findings have demonstrated that DNA methylation is a significant component in the hierarchy of control structures that regulate gene differentiation and function (Razin and Riggs 1980). Enhanced cytosine methylation at key sites is known to correlate with modified chromatin structure and gene expression (Flavell 1994). The importance of this epigenetic mechanism for normal development has been recorded both in plants and animals (Fojtová *et al* 2000).

Studies that have included plants, animals, and fungi all support the idea that gene silencing is a key biological consequence of DNA methylation (Bird and Wolffe 1999). A substantial amount of DNA methylation in vertebrates is present within genes; outside of vertebrates, where almost the entire genome is methylated, only a few methylated genes

have been described (Zilberman *et al* 2007), and the biological significance of genic DNA methylation is only beginning to be elucidated, mainly through genetic and genomic approaches used on *Arabidopsis thaliana* (Aceituno *et al* 2008, Inagaki and Kakutani 2010). Concerning plants, methylation occurs at symmetric CpXpG and asymmetric CpXpX sites, as well as CpG sites that are also recognized in the majority of mammalian genomes (Rabinowicz *et al* 2003).

DNA methyltransferases, which can be found in both plants and animals, maintain methylation of CpG dinucleotides (Rabinowicz *et al* 2003). Methylation of CpXpG and asymmetric sites is implemented by chromomethylases; these enzymes are encoded by plant-specific gene families (Rabinowicz *et al* 2003). DNA methylation can also be affected by chromatin-remodeling proteins (Rabinowicz *et al* 2003). Chromatin proteins can influence the accessibility of DNA to methylase and hence the cytosine methylation pattern (Flavell 1994). It is also known that the distribution of methylated cytosine can affect the complement of chromatin proteins, chromatin condensation, and accessibility of the chromatin to transcription complexes and specific transcription factors (Flavell 1994).

5mC makes up about 2-7% of the full amount of cytosine in animal cells, while over 25% of the cytosine residues are methylated in plant cells (Gruenbaum *et al* 1981). This dissimilarity is due to the fact that, though plants and animals both possess CG sequences methylated to about 70-80%, the CG dinucleotide is much more prevalent in plant DNA (3-4%) than it is in animal DNA (0.5-1%) (Gruenbaum *et al* 1981). The

elevated 5mC levels found in plants may also be attributed to methylation at other sites; this has been ascertained through partial sequence analysis and restriction enzyme analyses, both of which demonstrate that nearly all methylated CA, CC, and CT residues are located in the symmetrical trinucleotide sequence CXG (Gruenbaum *et al* 1981). Like that in vertebrate cells, the methylatable sites of plant DNA are not entirely methylated (Gruenbaum *et al* 1981). Roughly 80% of all CG or CAG/GTC sites are methylated, with no more than 50% of the CCG sites methylated at the external cytosine (Gruenbaum *et al* 1981). It is this incomplete methylation which makes those sites potential factors in cellular regulation (Gruenbaum *et al* 1981).

The majority of biological modifications such as protein phosphorylation can be reversed; there exist enzymes that can catalyze this process either through further modification or through removal of the modification itself (Ramchandani *et al* 1999). This reversibility is crucial in order to operate as biological signals that have the ability to react to changing physiological cues (Ramchandani *et al* 1999**).** DNA methylation was once thought to be an exception to this rule because in order to remove a methyl group from DNA a cleavage of a carbon–carbon bond must take place, which was considered an unlikely reaction (Ramchandani *et al* 1999). It is now accepted that demethylation must occur during development. It was originally proposed that this is accomplished through nondirect mechanisms involving base excision and repair for removal of methyl groups from DNA at various stages of development (Ramchandani *et al* 1999). Because it was believed that the removal of methyl groups from DNA is an energetically costly procedure, the conventional model has been that DNA methylation is both a heritable and

stable signal (Ramchandani *et al* 1999). Ramchandani et al identified a DNA demethylase that catalyzes the cleavage of a methyl residue from 5-methyl cytosine and provided molecular proof that it is biochemically possible to demethylate DNA (Ramchandani *et al* 1999). In addition, studies conducted on mouse embryonic fibroblasts (MEFs) by Dodge et al revealed that constitutive or conditional deletion of methyltransferase Dnmt3b resulted in partial loss of DNA methylation throughout the genome, suggesting that *Dnmt3b*, in addition to the major maintenance methyltransferase *Dnmt1*, is required for maintaining DNA methylation in MEF cells (Dodge *et al* 2005).

There still is much to be learned about the mechanism of DNA methylation in eukaryotes. It is known that the process of methylation is facilitated by a class of enzymes known as DNA methyltransferases. It has been suggested that there are 2 types of DNA methyltransferase activity: maintenance methylation and de novo methylation. Maintenance methylation takes place when the pattern of methylation on the parent DNA strand serves as the template for methylation on the newly synthesized daughter DNA strand **(**Genereux *et al* 2005). De novo methylation occurs when methyl groups are added at unmethylated dyads without regard to template pattern (Genereux *et al* 2005). Although stress-induced changes in euchromatic gene transcription are well recorded, there is only some degree of support so far that environmental stimuli can lessen the extreme suppression of transcription in heterochromatin, which appears to be constitutively silenced by layers of epigenetic control (Tittel-Elmer *et al* 2010). Radiation-induced genome instability has been proposed to be characteristically epigenetic (Koturbash *et al* 2006). For instance, in a study on *Arabidopsis thaliana*, it

was found that stress induced by UV-B exposure caused immediate and heritable activity of a previously silenced gene under epigenetic control (Lang-Mladek *et al* 2010*).* A plethora of DNA damaging agents, especially radiation, are understood to alter DNA methylation patterns in the genome which could contribute to their genome destabilizing effects (Koturbash *et al* 2006).

Novel methods have been utilized to reveal methylated sequences in specific regions of the chromosome. One such method was performed through the use of bacterial restriction endonucleases which cleave DNA at specific sites (Razin and Riggs 1980). The majority, around 90%, of the 5mC residues in eukaryotic DNA are found in the dinucleotide sequence CpG (Razin and Riggs 1980). Several restriction enzymes recognize CpG as their target sequence (Razin and Riggs 1980). Some of these "CpG enzymes" (and cutting sites) are the restriction endonucleases HpaII (CCGG), Msp I (CCGG), HhaI (GCGC), XhoI (CTCGAG), AvaI (CPyCGPuG), SalI (GTCGAC), and SmaI (CCCGGG) (T, thymine; A, adenine; Py, pyrimidine). Most of these enzymes are sensitive to methylation and will not cut at the CpG sequence site if it is methylated. Thus, these methylation-sensitive enzymes can be utilized to search for methylation sites (Razin and Riggs 1980).

In practice, for meaningful results, it is essential to have a control showing full digestion, so the pattern to be expected in the absence of methylation is known. A more convenient and appropriate method became feasible when it was discovered that MspI recognizes and cleaves the CCGG sequence in the same way as does HpaII. However,

MspI cuts the DNA whether or not the internal cytosine is methylated (Razin and Riggs 1980). Analyses in the past have utilized a methylation sensitive variant of the amplified fragment length polymorphism (AFLP) technique to help in identifying a subset of methylations in genomes of interest (Xiong *et al* 1999). This is a strong method for fingerprinting genomes that unites two major methods used in molecular analysis: restriction fragment length polymorphism (RFLP) and the polymerase chain reaction (PCR).

To summarize, the AFLP strategy entails the genomic digestion of DNA with two restriction enzymes. The sticky overhangs from the newly digested DNA are then ligated to double stranded adapter molecules of known sequence; these adapters are used as primer templates in the first round of PCR. Primers that are complementary both to the adapters and to the restriction sites are used to amplify the DNA fragments which are flanked by adapters. For the purposes of selective reduction of PCR amplified products, primers with additional nucleotide(s) at their 3' end are used. This ensures that only fragments that carry complementary nucleotides to these selective primers will be amplified. In methylation-sensitive AFLP, isoschizomers HpaII and MspI are used in separate, identical reactions. Both restriction enzymes target the sequence CCGG for digestion but differ in their sensitivity to DNA methylation at the inner cytosine (Jeddeloh *et al* 1996). When the resulting amplified fragments are run alongside one another on a gel, a difference in band pattern indicates a methylation polymorphism.

Two analyses to study methylation within the genomes of W. Murcott and Tango *Citrus* trees, LUMA pyrosequencing analysis and methylation-sensitive amplified fragment length polymorphism (MS-AFLP), were conducted to determine whether or not the percentage of methylated cytosines or the pattern of methylated cytosines within the respective *Citrus* genomes were altered as a result of exposure to stress-inducing gamma radiation. One possible underlying cause of the seedless phenotype seen in Tango could be a change in methylation of its cytosines, which in turn could activate previously silenced transposable elements or alter the transcription pattern genes, either of which could result in a seedless phenotype. The LUMA pyrosequencing and MS-AFLP analyses will both target methylated cytosines within the genome of each *Citrus* variety using methylation-sensitive and insensitive restriction enzyme combinations. The LUMA pyrosequencing method provides an overall estimate of the percentage of methylated CpG whereas MS-AFLP analysis can detect specific sequences that differ in methylation pattern.

Materials And Methods

Tissue collection

Citrus tissue was collected from the research collection fields at UCR. Two tissue samples per tree were collected and stored for DNA extraction. The tissues of interest being immature flower buds and young leaves (*Table 4.1*). Immature flower buds ranged in size from 5-10 mm in length. W. Murcott tissue was collected from a single tree in fields 1, 12, and 13; for a total of 3 immature flower bud and 3 young leaf tissue samples. Tango tissue was collected from a single tree in fields 1, 13, and 15; for a total of 3 immature flower bud and 3 young leaf tissue samples. Tissue samples were kept cold on ice before long-term storage; samples were briefly washed using deionized tap water to remove dirt and any surface contaminants. Female organs (stigma, style, and ovary) were removed from the immature flower buds using laboratory forceps and a razor blade to remove petals, sepals, anthers, and other parts of the immature flower bud. Female organs were collected for DNA extraction while the removed parts of the immature flower buds were discarded. Samples were then wrapped individually in aluminum foil, immersed in liquid N_2, then stored in a -80°C freezer until needed for DNA extraction.

DNA extraction

Frozen tissue (1 g) was ground to a fine powder in a mortar and pestle with liquid N_2 and collected in a 50-ml collection tube. The powder was kept on dry ice so that ground tissue would remain frozen until ready for DNA extraction. DNA was extracted using the

Qiagen DNeasy Plant Maxi Kit; the manual protocol for large plant DNA extractions included in the kit was followed as written, with the exception of step 17. In step 17, the final DNA elution step, eluate from step 16 was pipetted from its 50 ml collection tube and was directly applied to the column membrane where it was allowed to incubate for 5 min at room temperature before being centrifuged (3000-5000x g) to elute. The final step was altered from the protocol so as to increase DNA recovery. Final eluates were aliquoted into 2-ml microcentrifuge tubes. DNA concentration was measured for each sample using a Nanodrop spectrophotometer. DNA extractions were stored in the -20°C freezer until needed for further use.

LUMA Methylation Analysis

DNA samples were thawed and diluted with ddH_2O to a concentration of 1 μg/ml. Samples were stored in -20°C freezer overnight. Frozen DNA samples were shipped to EpigenDX for methylation analysis.

LUMA procedure from EpigenDX: "Two equal aliquots of a sample DNA are double digested in parallel. The A digest is carried out with *Hpa*II and *Eco*RI and the B digest is carried out with *Msp*I and *Eco*RI. *Hpa*II and *Msp*1 are isochizomers that digest at the recognition sequence CCGG. *Hpa*II is sensitive to CpG methylation and will only digest unmethylated sequence, while MspI is not sensitive to CpG methylation and will digest all the recognition sites. *Eco*RI cuts at GAATTC, leaving AATT overhangs and is not affected by methylation. Following enzyme digestion each reaction is subjected to a pyrosequencing reaction. When nucleotides are filled into the restriction overhangs, light is given off in proportion to the number of nucleotides incorporated. The signal obtained

from the fill-in of the A and T nucleotides in *Eco*RI digest in both reactions is a control for the amount of DNA in the reaction and is used to normalize the value for the signal obtained from the fill-in of the G and C nucleotides of the *Hpa*II and *Msp*I digests. By calculating the ratio of the normalized *Hpa*II and *Msp*I digests, the overall percent methylation of a sample of DNA can be determined."

Methylation data from LUMA analysis was sent to the Roose lab as an excel spreadsheet for statistical analysis.

Statistical Analysis of LUMA Data

Methylation data was analyzed using the Analysis Toolpak on Microsoft Excel. A two-sample t-test was used with the assumption of equal variances between samples. Samples were evaluated and compared as follows: (1) W. Murcott immature flower bud tissue was evaluated against Tango immature flower bud tissue; (2) W. Murcott young leaf tissue was evaluated against Tango young leaf tissue; (3) W. Murcott immature flower bud tissue was evaluated against W. Murcott young leaf tissue; (4) Tango immature flower bud tissue was evaluated against Tango young leaf tissue; (5) W. Murcott immature flower bud tissue and young leaf tissue were evaluated against Tango immature flower bud tissue and young leaf tissue; W. Murcott immature flower bud tissue and Tango immature flower bud tissue were evaluated against W. Murcott young leaf tissue and Tango young leaf tissue. Total methylation percentage differences were considered to be statistically significant if the t-test probability value was less than or equal to 0.05.

DNA extraction for MS-AFLP analysis

Young leaf tissue was collected by Lisa Mu. Young leaf tissue was taken from W. Murcott and Tango *Citrus* trees located in the agricultural research fields at the University of California in Riverside. Leaves were collected from W. Murcott trees (field 12) and from Tango trees (fields 1 and 15) (*Table 4.2*). Leaves were washed to remove surface contaminants and were stored at -80°C until needed for DNA extraction. Frozen young leaf tissue (100 mg) was removed from the -80°C freezer and was ground in liquid N_2 using a mortar and pestle, and transferred to a 1.5-ml microcentrifuge tube. Ground tissue was held on dry ice until ready for DNA extraction. DNA extraction was carried out according to Qiagen DNeasy Plant Mini Kit instructions. DNA concentration and quality was measured for each sample using a Nanodrop spectrophotometer. DNAs were stored in the -20°C freezer.

Samples were each given an ID name along with description for each (*Table 4.2*) for this analysis. Adapters and primers used to prepare AFLP templates and for the polymerase chain reaction (PCR) component of MS-AFLP were also given an identification name, description, and full nucleotide sequence (*Table 4.3*).

Restriction Digestion

Restriction digestion of the extracted genomic DNA samples was performed by Lisa Mu and conducted as follows: three µl of 5X reaction buffer (from AFLP kit) was mixed with 9 µl of genomic DNA (180 ng) digested with 5 units of *Eco*RI/*Msp*I or *Eco*RI/*Hpa*II, in

1X AFLP reaction buffer in a total volume of 15 µl. The 180 ng of DNA was allowed to digest for 2 h at 37°C. Digested genomic DNAs (60 ng) and undigested DNAs (100 ng) were fractionated on a 1.4% agarose gel to test whether digestion of genomic DNA was complete. The 100-bp DNA ladder (Invitrogen) and 1 kb ladders (Invitrogen) were loaded into separate lanes to help measure fragment sizes and the extent of DNA migration through the gel. The gel was run at 100V for 1.5 hrs.

Ligation

Ligation of digested DNA to adaptors was performd by Lisa Mu. For ligation, all digested DNA underwent the same ligation reaction as follows: the *Hpa*II/*Msp*I adapter was formed by annealing equimolar amounts of the two DNA strands, mixing the strands together at 95°C for 10 min and then cooling to room temperature. Digested DNAs were ligated to adapters (*Table 4.*3) in a reaction mixture containing 0.5 µl 50 µM *Hpa*II/*Msp*I adapter, 0.5 µl 5 µM *Eco*RI adapter, 0.5 µl 10mM ATP, 1 µl 5X Restriction-Ligation buffer, 0.1 µl T4 DNA ligase (0.6 unit of NEB enzyme), 5-10 µl restriction digest product, volume to 10-15 µl with ddH_2O. The reaction was incubated at room temperature (20°C) for 2 h, then stored at 4°C overnight. The product was diluted to 1:10 with ddH_2O for preselective amplification.

Pre-selective Amplification

Pre-selective amplification was performed by Lisa Mu. After ligation, the restricted digested/ligated DNA samples were subjected to a preselective amplification using e-prea and hm-prea primer sets (*Table 4.3*) for the *Eco*RI and *Msp*I/*Hpa*II adapter sequence regions, respectively. The ligation mixture was diluted 1:10 and was used in the pre-

selective amplification as follows: 5 µl of 1 µM *Eco*RI-preamp primer, 5 µl of 1 µM *Hpa*II/*Msp*I-preamp primer, 1 µl 10X PCR $MgCl_2$ free buffer, 1 µl $MgCl_2$ (25mM), 0.2 µl Taq (5U/µl) DNA polymerase, 1.6 µl dNTPs (2mM), 2 µl DNA ligation product (1:10 diluted), ddH_2O used to volume reaction to 16.4 µl.

Polymerase chain reaction was performed for 21 cycles with 30 sec denaturation at 94°C, 1 min annealing at 56°C, and 1 min extension at 72°C. PCR products (5 µl) were fractionated on a 1.4% agarose gel with EtBr. The PCR product was diluted 1:20 with ddH_2O and was used as a DNA template for selective amplification.

MS-AFLP Selective PCR Reaction

DNA samples from the preselective PCR reaction were used in a selective PCR reaction, performed by Lisa Mu. Selective primer sets shared the core sequence of the amplified DNA adapters but contained two additional, 5' nucleotides to provide specificity (*Table 4.3*). A multiplexed PCR was used in which two different *Eco*RI primers labeled with different IR-fluorescent dyes (700nm and 800nm) were combined with a single *Hpa*II./*Msp*I primer. Template DNA (1:20 dilution preselective PCR products) was used in the following reaction mixture: 0.5 µl 700 *Eco*RI primer (1 µM), 0.5 µl 800 *Eco*RI primer (1 µM), 2 µl *Hpa*II./*Msp*I primer (1 µM), 1.2 µl dNTPs (2mM), 1.23 µl 10X PCR $MgCl_2$ free buffer, 0.74 µl $MgCl_2$ (25 mM), 0.14 µl Taq DNA polymerase (Promega), 2.32 µl diluted pre-amp DNA (1:20), volume to 12.33 µl with ddH2O.

MS-AFLP selective PCR reactions generally employed two oligonucleotide primers, one corresponding to the *Eco*RI-ends and one corresponding to the *Hpa*II/*Msp*I-ends. The

*Eco*RI-primer was fluorescence infrared-labeled (700 or 800). These primers targeting different pairs of nucleotides (*Table 4.3*) were mixed and matched together to create various combinations of *Eco*RI-site-specific and *Msp*I-*Hpa*II-site-specific primer sets. Polymerase chain reaction was performed under the following conditions: initial heat denaturing step of 94°C for 15 sec, followed by an annealing step of 65°C for 30 sec (-0.7 degrees each subsequent cycle), and an elongation step of 72°C for 1 min, followed by a repeat of denaturing, annealing, and elongation steps 12X; followed by 29 cycles of a denaturing step of 94°C for 15 sec, an annealing step of 56°C for 30 sec (+1 second each subsequent cycle), and an elongation step at 72°C for 1 min (+1 second each subsequent cycle),; followed by an elongation step of 72°C for 15 min, followed by an indefinite incubation at 4°C.

Polyacrylamide Gel Analysis

A 7% 25 cm polyacrylamide gel was prepared by Lisa Mu to analyze DNA samples from the selective PCR reaction. The gel was made with 20 ml of acrylamide gel stock (from 4 °C stored stock solution of: 63.0 g urea, 21.0 ml Long Ranger solution, 18.0 ml 10X Tris-Boric Acid EDTA (TBE) buffer; brought to 150 ml with ddH_2O), 15 µl TEMED, and 150 µl 10% ammonium persulfate (APS). The gel was allowed to polymerize at room temperature for about 30 min. After polymerization, a 65-well comb was inserted into the gel. The gel was pre-heated in a LI-COR model 4000L machine during a pre-run for 25 min in 1X TBE buffer before the DNA samples were loaded.

Five µl of LI-COR loading dye (from stock: 47.5 ml formamide, 2 ml 0.5M EDTA, 0.5 ml ddH_2O, 40 mg bromophenol blue) was added to each DNA sample from the selective

PCR reaction. DNAs were denatured using a thermocycler programmed to run as follows: 94°C for 3 min 30 sec, followed by an indefinite cycle of 4°C. After denaturation, DNAs were loaded onto the polyacrylamide gel (about 1-2 µl per well depending on primer set) along with 700-bp DNA ladder (from LiCor). Gels were run for 2-3 hours.

Analysis of Li-Cor gel images

Li-Cor gel images were collected from each *Eco*RI/*Hpa*II-*Msp*I primer set combination (700nm and 800nm images were generated from the Li-Cor software system). Banding patterns generated from *Eco*RI/*Msp*I-digested and *Eco*RI/*Hpa*II-digested DNAs were compared. Each primer set combination for each DNA digest was scored for the number of bands that were the same for each digest, unique to one digest, or more prominent in one digest than in the other. The DNA ladder was used as a reference when sizing the bands to ensure that only bands of the same size were compared with one another. If the same sized bands were present in both DNA digests at the same intensity, they were scored as "same". If certain sized bands were present in one DNA digest but absent in the other, the bands were scored as "unique" to the digest containing them. If certain sized bands were present in both DNA digests but had a stronger intensity in one of the digests, the bands were scored as "prominent" in the digest containing the stronger bands. The number of scoreable bands was tabulated for each primer combination and each band was classified as same (shared in presence and intensity by both types of digests), unique (present in only one digest), or prominent (more intense in one digest). Scored bands

were tallied for each category; these numbers were used in statistical analysis of methylated-cytosine patterns.

Results

Data and results from LUMA pyrosequencing analysis

DNA samples were extracted from immature flower bud and young leaf tissue taken from one W. Murcott or Tango tree from one of several agricultural research fields at the University of California in Riverside. Samples from three W. Murcott and three Tango trees in four agricultural fields were each given an ID name along with description for each (*Table 4.1*) and submitted to EpiGenDX for LUMA pyrosequencing analysis. Data generated from the LUMA pyrosequencing analysis was generated in replicate for each sample submitted (*Appendix C*). Duplicate samples were analyzed to determine the percentage of methylated cytosines within its genome (*Appendix C*). The mean of these values was calculated for each sample to determine the average percentage of methylated cytosines (*Appendix C*). From the results, one outlier (WM-yl-12-48-4) was detected in one of the W. Murcott samples (*Appendix C*) which showed an elevated level of methylated cytosines when compared to its replicate as well as the other *Citrus* DNA samples. EpigenDx staff agreed that this value was likely to reflect a technical problem with this particular measurement. For these reasons, the outlier was discarded when conducting the statistical analysis on the data; only its replicate, which appeared to not have such a skewed value of methylation percentage, was included in the analysis (*Appendix C*).

A total of six comparisons were made: between genotypes for each tissue type and overall, and between tissues for each genotype and overall (*Table 4.4*). The variances

of all treatment combinations were homogeneous using the F-test (*Table 4.4*), allowing Student's t-test with equal variances to be used to test for differences between groups. The cutoff P-value for a significant difference was set at 0.05.

The percentage of methylated cytosines ranged from 53.1% in Tango flower buds to 58.6% in W. Murcott leaf tissue. With the exception of W. Murcott and Tango immature flower bud DNA vs. W. Murcott and Tango young leaf DNA, all other comparisons were found to have a P-value higher than 0.05 (*Table 4.4*). These results indicate that there was no significant difference in average percentage of methylated cytosines between Tango and W. Murcott in either young leaves or immature flowers. However, when averaged over genotypes, immature flower bud DNA had significantly less methylated cytosine (53.3%) than young leaf DNA (57.2%).

Data and results from MS-AFLP analysis (performed by Lisa Mu)

DNA samples digested by the *Eco*RI/*Msp*I enzyme combination and by the *Eco*RI/*Hpa*II enzyme combination were analyzed for complete digestion and compared with undigested control DNA samples. The undigested control samples showed a distinct single band of a large size (>10 kb), whereas the digested DNA samples showed a long smear running down the length of the gel (*Figure 4.1*), as expected for samples digested to completion.

After performing the preselective PCR reaction on digested, adapter-ligated genomic DNAs, analysis of the PCR products on an agarose gel confirmed that the preselective PCR reaction had worked (*Figure 4.2*). Similar banding patterns were

observed among DNA samples that had been digested by the same restriction enzyme combinations.

The selective PCR products were analyzed on polyacrylamide gels run on the LiCor system (*Figures 4.3, 4.4)*. Banding patterns were compared between W. Murcott and Tango samples 1-6 (*Table 4.2*). Results indicated that, for each group of six samples that were amplified using a particular primer set combination, there was no difference in the banding patterns (*Figures 4.3, 4.4*). There were differences in banding patterns between groups of samples amplified by the same selective primer set but digested by a different set of enzymes; these differences in banding were scored to estimate the number of CpG methylation sites analyzed in this study.

The numbers of scored bands in each descriptive category for each DNA digest were summed to estimate the total scored bands in each category (*Tables 4.5, 4.6*). The percentage of total bands analyzed for each category in each DNA digest was calculated (*Table 4.7*) based on the sum of all bands analyzed for that category divided by the total number of bands that were analyzed. The percentages of both unique and prominent scoreable bands found in the *Eco*RI/*Msp*I digest were higher than those found in the *Eco*RI/*Hpa*II digest. This is not surprising when one considers that *Msp*I is not inhibited by methylation to the same degree as *Hpa*II and therefore would be expected to generate more digestion products. The total number of target sites analyzed within the *Citrus* genome was estimated as 1457 (*Table 4.6*) based on the total number of bands shared by both types of digests (assumed to be cutting at the same target sites) added to the total

number of unique bands found in the *Eco*RI/*Msp*I digest (assumed to be methylated target sites) added to the total number of prominent bands found in *Eco*RI/*Msp*I and *Eco*RI/*Hpa*II digests. The total number of unique bands found in the *Eco*RI/*Hpa*II digest was excluded from this sum as it was assumed that these target sites were already accounted for in the unique bands found in the *Eco*RI/*Msp*I digest. An estimate of the total number of methylated sites that were analyzed was calculated (*Table 4.7*) by adding the total number of unique bands found in the *Eco*RI/*Msp*I digest to the total number of prominent bands found in the *Eco*RI/*Msp*I digest (these are assumed to be partially methylated target sites). An estimate of the total number of unmethylated sites that were analyzed was calculated (*Table 4.7*) by adding the total number of bands shared by both types of digests to the total number of prominent bands found in the *Eco*RI/*Hpa*II digest. The percentage of methylated sites was estimated at 52%. Of the total sites that were analyzed, 8.6% were considered to be partially methylated based on bands that were more prominent in, but not unique to, the *Eco*RI/*Msp*I digested DNA. The percentage of unmethylated sites was estimated at 48%, which consisted of 576 bands that were scored as identical between both *Eco*RI/*Msp*I and *Eco*RI/*Hpa*II digests (26.4%) and 124 bands that were scored as more prominent in, but not unique to, the *Eco*RI/*Hpa*II digested DNA (5.7%; these are assumed to be unmethylated sites that accumulated in greater amounts in the *Eco*RI/*Hpa*II digest).

Discussion

Based on results obtained both from LUMA pyrosequencing analysis and from MS-AFLP analysis, no significant differences in levels of genomic methylation were detected between W. Murcott and Tango. It is important to note that both LUMA pyrosequencing analysis and MS-AFLP analysis rely on the usage of methylation-sensitive restriction enzymes. One drawback of these technologies is that all restriction enzyme-based techniques are limited to analysis of methylation within recognition sites. MS-AFLP analysis failed to detect polymorphic fragments between W. Murcott and Tango DNA samples, indicating that the DNA methylation state remains unaltered in the genomic regions that were analyzed.

There were a total of 1,457 genomic sites analyzed in MS-AFLP, 757 of which are estimated to be methylated at the inner cytosine of the recognition sequence CCGG. This equates to about 52% of the CCGG sites analyzed as being methylated at the inner cytosine; only sites at which the inner cytosine has been methylated can be measured, since *Msp*I cleaves C^{5m}CGG but not ^{5m}CCGG sequences (McClelland *et al* 1994). Statistical analysis performed on the average percentage of genomic methylation generated from LUMA pyrosequencing analysis revealed that the probability of methylation percentage averages being similar between W. Murcott and Tango populations was high, thus suggesting that overall genomic methylation levels found in the W. Murcott and Tango populations and tissues studied are similar to one another.

The two approaches to detect differences in DNA methylation used here are complementary in that LUMA analysis estimates the genomewide percentage of CCGG restriction sites that are methylated. MS-AFLP analysis characterizes the methylation state of a large number of individual sites which together comprise only a small fraction of the genome. Thus if 1% of the targeted C sites change in methylation status, this would be detected by MS-AFLP analysis, but not by LUMA analysis.

According to Koturbash et al, DNA polymerases involved in repair and recombination will include cytosine but not methylated cytosine during repair synthesis. Thus, the generation of DNA lesions by such DNA breakage-inducing agents as gamma irradiation and the subsequent activation of DNA repair and recombination mechanisms may result in DNA hypomethylation (Koturbash *et al* 2006). In the study conducted by Pogribny *et al*, three *in vivo* murine model groups were exposed to three radiation doses over a period of 10 days (whole-body irradiation of 0.05 Gy X-ray daily for 10 consecutive days for a cumulative dose of 0.5 Gy in fractionated radiation group, 0.5 Gy X-ray only once in the acute radiation group, 0 Gy X-ray in the control group) and the epigenetic changes in thymus in male and female mice was determined using a sensitive *Hpa*II-based cytosine extension assay that measured the proportion of unmethylated CCGG sites (Pogribny *et al* 2005). It was discovered that acute application of 0.5 Gy X-ray led to a 1.8- and 2.3-fold ($P < 0.05$) increase in the [^{3}H]dCTP incorporation in female and male thymus, respectively, indicative of hypomethylation of DNA, while fractionated low-dose radiation exposure resulted in even more substantial decrease in global DNA

methylation than acute irradiation, causing 2.5- and 6.1-fold ($P < 0.05$) reduction in global DNA methylation in females and males, respectively.

A decrease in global methylation in irradiated mouse skin correlated with the accumulation of DNA strand breaks as well as with the increase in recombination activity, suggesting that the loss of methylation in cells from directly irradiated tissue may be related to the repair status of these cells (Koturbash *et al* 2006). Indeed, Kovalchuk *et al* found in their studies that mice exposed to an acute dose of radiation (7.5 Gy) displayed significantly increased average frequencies of recombinant cells by from 1.1 to 15.1 per 105 cells as measured by flow cytometry (Kovalchuk *et al* 2004). As hypomethylation could not be detected in Tango DNA when compared to W. Murcott DNA, it cannot be concluded with certainty that exposure to gamma radiation resulted in DNA breakage in methylated cytosine regions of DNA. However, it is possible that small amounts of DNA breakage may have occurred, thereby limiting the amount of hypomethylation present in the genome as to be undetectable by LUMA analysis. It is also possible that irradiation induced only a transient change in methylation in citrus, a change that has been largely or completely reversed in the 14 years and many cell generations since the tissue that became Tango was irradiated.

Aberrant cytosine DNA methylation is linked to genomic instability and increased rates of genome rearrangements (Koturbash *et al* 2006). A variety of DNA damaging agents including irradiation are known to affect genome DNA methylation patterns and this may contribute to their genome destabilizing effects (Koturbash *et al* 2006).

According to Ling et al, presently, there is no major report stating the use of gamma irradiation as a physical mutagen to alter the physiological characteristics of *C. sinensis* (Ling *et al* 2008). Surveys of the literature in regard to radiation treatment show that there is presently no consistent report regarding the effect of irradiation on the nucleic acids. Some investigators have observed slight depression or increase in protein synthesis, while others reported no significant changes (Ling *et al* 2008).

An interesting study was conducted by Lang-Mladek et al in which immediate and heritable effects of abiotic stresses, such as high and low-temperature stress and UV-B irradiation, on the epigenetic regulation of a transcriptionally silent GUS-labeled transgene locus in *Arabidopsis thaliana* were addressed (Lang-Mladek *et al* 2010). It was discovered that in the stressed generation, the silent TS–GUS transgene as well as endogenous silent transposons were transcriptionally reactivated (Lang-Mladek *et al* 2010). However, this stress-mediated release of gene silencing correlated with pronounced alterations in histone occupancy and in histone H3 acetylation but did not involve adjustments in DNA methylation (Lang-Mladek *et al* 2010). In addition, although strong with respect to its initial amplitude, the reactivated state persisted only transiently (length of time not given) and was eventually reversed in tissue formed after stress treatments in the recovery phase (Lang-Mladek *et al* 2010). Based on these findings, it is possible that although no significant differences in methylation patterns or in overall percentage of methylated cytosines within the genome could be found between W. Murcott and Tango, other epigenetic alterations may have occurred within Tango in response to stress-inducing gamma radiation treatment which may have activated

previously silenced genes, thus causing a seedless phenotype. It would be worthwhile to look into other epigenetic alteration possibilities, such as histone modification, to detect any possible significant differences that may exist between W. Murcott and Tango.

One interesting finding from the LUMA pyrosequencing analysis was that there was a significant difference in percentage of methylated cytosine within the genome when comparing DNA extracted from immature flower buds with DNA extracted from young leaves in *Citrus*. According to findings by Zhang et al, the relationship between DNA methylation and gene expression patterns reveals a correlation between the position of DNA methylation within a gene, the level of gene expression, and its tissue specificity (Zhang *et al* 2006). It was found that a distinct fraction of promoter-methylated genes have a dramatically higher tissue specificity, suggesting that promoter-methylated genes tend to be expressed in a tissue-specific manner (Zhang *et al* 2006). According to data acquired from LUMA pyrosequencing analysis, the percentage of methylated cytosines was significantly lower in immature flower bud tissue than in young leaf tissue. The lower level of methylation in immature flower buds may indicate that more of the chromatin is available for transcription. The percentage of estimated methylated target sites using the methylation-sensitive AFLP analysis (52%) was close to the percentage of unmethylated target sites (48%); this result is somewhat lower than the percentage of methylation in the Citrus genome estimated using LUMA analysis (about 55-59%). However, the methods used to estimate methylated target sites in ms-AFLP analysis were applied to only a small portion of all potential methylated sites and therefore are subject to sampling error. LUMA has the advantage of computerized measurements that are able

to analyze an entire genome, therefore, an estimate of 52% methylation is acceptable when accounting for inaccuracy that comes with sampling error. Further experimentation could be conducted to identify the specific location of the unmethylated cytosines; it would be of great interest to find out whether or not these cytosines are located in promoter regions and, if so, which genes are linked to these promoters.

References

Aceituno F.F., Moseyko N., Rhee S.Y., and Gutierrez R.A., 2008. The rules of gene expression in plants: organ identity and gene body methylation are key factors for regulation of gene expression in *Arabidopsis thaliana*. BMC Genomics 9:438–451.

Bird A.P., and Wolffe A.P., 1999. Methylation-induced repression--belts, braces, and chromatin. Cell 99: 451–454.

Cheng C.,Daigen M.,and Hirochika H., 2006. Epigenetic regulation of the rice retrotransposon *Tos17*. Mol. Gen. Genomics 276:378–390.

de Assis Alves Mourao Filho F., Pio R., Mendes B.M.J., de Azevedo F.A., Schinor E.H., Entelmann F.A., Alves A.S.R., and Cantuarias-Aviles T.A., 2008. Evaluation of citrus somatic hybrids for tolerance to *Phtophthora nicotianae* and citrus tristeza virus. Scientia Horticulturae 115: 301-308.

Dodge J.E., Okano M., Dick F.,Tsujimoto N., Chen T., Wang S., Ueda Y., Dyson N., and Li E., 2005. Inactivation of *Dnmt3b* in mouse embryonic fibroblasts results in DNA hypomethylation, chromosomal instability, and spontaneous immortalization. J. Biol. Chem. 280: 17986-17991.

Flavell R.B., 1994. Inactivation of gene expression in plants as a consequence of specific sequence duplication. Proc. Natl. Acad. Sci. USA 91: 3490-3496.

Fojtova´ M. Kovarˇı´k A., and Matya´sˇek R., 2000. Cytosine methylation of plastid genome in higher plants. Fact or artefact? Plant Science 160: 585–593.

Genereux D.P., Miner B.E., Bergstrom C.T., and Laird C.D., 2005. A population-epigenetic model to infer site-specific methylation rates from double-stranded DNA methylation patterns. Proc. Natl. Acad. Sci. USA 102: 5802–5807.

Gruenbaum Y., Naveh-Many T., Cedar H., and Razin A., 1981. Sequence specificity of methylation in higher plant DNA. Nature 292: 860-862.

Hollick J.B., and Springer N., 2009. Epigenetic phenomena and epigenomics in maize. Epigenomics 10: 119-147.

Jeddeloh J. A., and Richards E.J., 1996. mCCG methylation in angiosperms. The Plant Journal. 9: 579-586.

JinPing X., LiGeng C., Ming X., HaiLin L., and WeiQi Y., 2009. Identification of AFLP fragments linked to seedlessness in Ponkan mandarin (*Citrus reticulata* Blanco) and conversion to SCAR markers. Scientia Horticulturae 121: 505-510.

Kamiri M., Stift M., Srairi I., Costantino G., Moussadik A.E., Hmyene A., Bakry F., Ollitrault P., and Froelicher Y., 2011. Evidence for non-disomic inheritance in a Citrus interspecific tetraploid somatic hybrid between *C. reticulata* and *C. limon* using SSR markers and cytogenetic analysis. Plant Cell Rep. 30:1415–1425.

Koturbash I., Rugo R.E., Hendricks C.A., Loree J., Thibault B., Kutanzi K., Pogribny I., Yanch J.C., Engelward B.P., and Kovalchuk O., 2006. Irradiation induces DNA damage and modulates epigenetic effectors indistant bystander tissue *in vivo*. Oncogene 25: 4267-4275.

Kovalchuk O., Hendricks C.A., Cassie S., Engelward A.J., and Engelward B.P., 2004. In vivo recombination after chronic damage exposure falls to below spontaneous levels in "recombomice". Mol. Cancer Res. 2: 567.

Lang-Mladek C., Popovab O., Kiokb K., Berlingera M., Rakicb B., Aufsatzb W., Jonakb C., Hausera M.T., and Luschniga C., 2010. Transgenerational inheritance and resetting of stress-induced loss of epigenetic gene silencing in *Arabidopsis*. Molecular Plant 3: 594–602.

Ling A.P.K., Chia J.Y., Hussein S., and Harun A.R., 2008. Physiological responses of *Citrus sinensis* to gamma irradiation. World Applied Sciences Journal 5: 12-19.

McClelland M., Nelson M., and Raschke E., 1994. Effect of site-specific modification on restriction endonucleases and DNA modification methyltransferases. Nucleic Acids Res. 22: 3640–3659.

McClintock B., 1951 Chromosome organization and genic expression. Cold Spring Harb Symp Quant Biol 16: 13–47.

Nagar S., Smith L.E., and Morgan W.F., 2003. Characterization of a novel epigenetic effect of ionizing radiation: the death-inducing effect. Cancer Res. 63:324–328.

Pogribny I., Koturbash I., Tryndyak V., Hudson D., Stevenson S.M., Sedelnikova O., Bonner W., and Kovalchuk O., 2005. Fractionated low-dose radiation exposure leads to accumulation of DNA damage and profound alterations in DNA and histone methylation in the murine thymus. Mol. Cancer Res. 3: 553-561.

Puig M., Cáceres M., and Ruiz A., 2004. Silencing of a gene adjacent to the breakpoint of a widespread Drosophila inversion by a transposon-induced antisense RNA. Proc. Natl. Acad. Sci. USA 101: 9013-9018.

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

Ramchandani S., Bhattacharya S.K., Cervoni N., and Szyf M., 1999. DNA methylation is a reversible biological signal. Proc. Natl. Acad. Sci. 96: 6107–6112.

Razin A., and Riggs A.D.,1980. DNA Methylation and gene function. Science 210: 604-610.

Inagaki S., and Kakutani T., 2010. Control of genic DNA methylation in Arabidopsis. J. Plant Res.123: 299–302.

Sutarto I., Agisimanto D., and Supriyanto A., 2009. Development of promising seedless *Citrus* mutants through gamma irradiation. Induced Plant Mutations in the Genomics Era. Food and Agriculture Organization of the United Nations, Rome: 306-308.

Tittel-Elmer M., Bucher E., Broger L., Mathieu O., Paszkowski J., Vaillant I., 2010. Stress-induced activation of heterochromatic transcription. PLoS Genetics 6: 1-11.

Vardi A., Levin I., and Carmi N., 2008. Induction of seedlessness in citrus: from classical techniques to emerging biotechnological approaches. J. Amer. Soc. Hort. Sci. 133 (1): 117-126.

Xiong L.Z., Xu C.G., Saghai Maroof M.A., and Zhang Q., 1999. Patterns of cytosine methylation in an elite rice hybrid and its parental lines, detected by a methylation sensitive amplification polymorphism technique. Mol. Gen. Genet. 261:439-446.

Zhang X, Yazaki J., Sundaresan A., Cokus S., Chan S.W.L., Chen H., Henderson I.R., Shinn P., Pellegrini M., Jacobsen S.E., and Ecker J.R., 2006. Genome-wide high-resolution mapping and functional analysis of DNA methylation in *Arabidopsis*. Cell 126: 1189–1201.

Zilberman D., Gehring M.,Tran R.K., Ballinger T., and Henikoff S., 2007. Genome-wide analysis of *Arabidopsis thaliana* DNA methylation uncovers an interdependence between methylation and transcription. Nature Genetics 39: 61-69.

Table 4.1:W. Murcott and Tango DNA Samples Used in LUMA Pyrosequencing Analysis

Sample ID	Source	Description
Tan-fb-15-2-10	UCR agricultural research field 15f, row 2, tree 10	Tango immature flower bud tissue
Tan-yl-13-34-11	UCR agricultural research field 13, row 34, tree 11	Tango young leaf tissue
WM-yl-1-20-13	UCR agricultural research field 1, row 20, tree 13	W. Murcott young leaf tissue
WM-fb-1-20-13	UCR agricultural research field 1, row 20, tree 13	W. Murcott immature flower bud tissue
WM-fb-13-14-14	UCR agricultural research field 13, row 14, tree 14	W. Murcott immature flower bud tissue
Tan-fb-1-28-3	UCR agricultural research field 1, row 28, tree 3	Tango immature flower bud tissue
Tan-yl-1-28-3	UCR agricultural research field 1, row 28, tree 3	Tango young leaf tissue
WM-fb-12-48-3	UCR agricultural research field 12, row 48, tree 3	W. Murcott immature flower bud tissue
WM-yl-13-14-14	UCR agricultural research field 13, row 14, tree 14	W. Murcott young leaf tissue
Tan-fb-13-34-11	UCR agricultural research field 13, row 34, tree 11	Tango immature flower bud tissue
Tan-yl-15-2-10	UCR agricultural research field 15, row 2, tree 10	Tango young leaf tissue
WM-yl-12-48-4	UCR agricultural research field 12, row 48, tree 4	W. Murcott young leaf tissue

Table 4.2: W. Murcott and Tango DNA Samples Used in MS-AFLP Analysis

Sample ID	Description
Tango-1	Seedless Tango mandarin field 15
Tango-2	Seedless Tango mandarin field 1
W. Murcott-3	Seedy W. Murcott mandarin field 12
W. Murcott-4	Seedy W. Murcott mandarin field 12
DNA #2	Seedless Tango mandarin
DNA #13	Seedy W. Murcott mandarin

Table 4.3: Primer Sequences Used in MS-AFLP Analysis of W.Murcott/Tango

Number	Name	Sequence 5'-3'	Length	Note
1	e-olig-1	CTCGTAGACTGCGTACC	17	*Eco*RI adapter 1
2	e-olig-2	AATTGGTACGCAGTC	15	*Eco*RI adapter 2
3	e-prea	GACTGCGTACCAATTCA	17	*Eco*RI prea primer
4	E-AAC	GACTGCGTACCAATTCAAC	19	selective *Eco*RI primer
5	E-AAG	GACTGCGTACCAATTCAAG	19	selective *Eco*RI primer
6	E-ACT	GACTGCGTACCAATTCACT	19	selective *Eco*RI primer
7	E-ACG	GACTGCGTACCAATTCACG	19	selective *Eco*RI primer
8	E-AAG	GACTGCGTACCAATTCAAG	19	selective *Eco*RI primer
9	E-ACA	GACTGCGTACCAATTCACA	19	selective *Eco*RI primer
10	E-AGC	GACTGCGTACCAATTCAGC	19	selective *Eco*RI primer
11	E-ACC	GACTGCGTACCAATTCACC	19	selective *Eco*RI primer
12	E-AGG	GACTGCGTACCAATTCAGG	19	selective *Eco*RI primer
13	h-m-olig-1	GACGATGAGTCTCGAT	16	*Hpa*II/*Msp*I adapter 1
14	h-m-olig-2	CGATCGAGACTCAT	14	*Hpa*II/*Msp*I adapter 2
15	hm-prea	ATGAGTCTCGATCGGA	16	*Hpa*II/*Msp*I prea primer
16	Hm1	ATGAGTCTCGATCGGAAT	18	selective *Hpa*II/*Msp*I primer
17	Hm2	ATGAGTCTCGATCGGATC	18	selective *Hpa*II/*Msp*I primer
18	Hm3	ATGAGTCTCGATCGGACT	18	selective *Hpa*II/*Msp*I primer
19	Hm4	ATGAGTCTCGATCGGAAA	18	selective *Hpa*II/*Msp*I primer
20	Hm5	ATGAGTCTCGATCGGATT	18	selective *Hpa*II/*Msp*I primer
21	Hm6	ATGAGTCTCGATCGGACG	18	selective *Hpa*II/*Msp*I primer

Primer and adapter sequences used in the MS-AFLP analysis of W. Murcott and Tango are shown along with a short description of each; note that selective primers are identical in sequence with one another with the exception of the final two base pairs

Table 4.4: Analysis Of LUMA Tests For Methylation in Mandarin DNA

Comparison		Mean ± S.D.	Mean ± S.D.	F-test	P (for F test)	t-test	P(for t-test)
Tango vs W. Murcott		Tango	W. Murcott				
	Flower buds	53.1 ± 1.74	53.4 ± 4.08	0.18	0.15	0.09	0.93
	Young leaf	55.9 ± 1.99	58.6 ± 1.01	3.87	0.21	2.02	0.11
	Combined	54.5 ± 2.27	56.0 ± 3.89	2.92	0.13	0.77	0.46
Flower vs leaf		Flower buds	Young leaf				
	Tango	53.1 ± 1.74	55.9 ± 1.99	0.77	0.43	1.84	0.14
	W. Murcott	53.4 ± 4.08	58.6 ± 1.01	16.21	0.06	2.14	0.10
	Combined	53.3 ± 2.81	57.2 ± 2.01	1.95	0.24	2.84	0.02

Mean and standard deviation of each comparison, F-test for heterogeneity of variances, and t-test for equality of means are shown for each comparison

Table 4.5: Number of Bands Scored as Identical or Different Between EcoRI/MspI and EcoRI/HpaII Primer Sets Analyzed on 700nm Channel

Primerset	Identical	EcoRI/MspI digested unique banding	EcoRI/HpaII digested unique banding	EcoRI/MspI digested prominent banding	EcoRI/HpaII digested prominent banding	Total scored bands
E-AAG/Hm1	22	11	1	4	6	44
E-AAG/Hm2	12	7	6	4	9	38
E-AAG/Hm3	11	13	0	4	0	28
E-AAG/Hm4	24	12	0	6	3	45
E-AAG/Hm5	0	15	5	1	3	24
E-AAG/Hm6	29	3	0	3	0	35
E-AAC/Hm1	9	1	0	2	4	16
E-AAC/Hm2	9	9	0	4	1	23
E-AAC/Hm3	14	2	1	1	0	18
E-AAC/Hm4	22	3	1	1	1	28
E-AAC/Hm5	10	7	0	1	2	20
E-ACA/Hm1	0	6	6	0	1	13
E-ACA/Hm2	0	3	8	1	1	13
E-ACA/Hm3	1	16	5	0	0	22
E-ACA/Hm4	27	8	3	2	2	42
E-ACA/Hm5	14	10	6	3	14	47
E-ACA/Hm6	12	12	1	0	1	26
E-ACC/Hm1	25	8	0	9	1	43
E-ACC/Hm2	18	16	0	9	5	48
E-ACC/Hm3	5	26	9	1	6	47
E-ACC/Hm4	0	14	9	0	3	26
E-ACC/Hm5	3	36	4	13	2	58
E-ACC/Hm6	8	19	12	10	9	58

Table 4.6: Number of Bands Scored as Identical or Different Between EcoRI/MspI and EcoRI/HpaII Primer Sets Analyzed on 800nm Channel

Primerset	Identical	EcoRI/MspI digested unique banding	EcoRI/HpaII digested unique banding	EcoRI/MspI digested prominent banding	EcoRI/HpaII digested prominent banding	Total scored bands
E-ACT/Hm1	14	4	0	1	0	19
E-ACT/Hm2	14	11	0	3	0	28
E-ACT/Hm3	5	4	1	1	2	13
E-ACT/Hm4	19	10	0	3	1	33
E-ACT/Hm5	7	20	0	10	4	41
E-ACT/Hm6	23	6	1	1	1	32
E-ACG/Hm1	42	15	1	12	2	72
E-ACG/Hm2	20	12	1	11	1	45
E-ACG/Hm3	20	18	0	11	0	49
E-ACG/Hm4	46	12	1	11	1	71
E-AGG/Hm1	5	11	1	8	0	25
E-AGG/Hm2	2	15	0	4	1	22
E-AGG/Hm3	5	26	9	1	6	47
E-AGG/Hm4	0	14	9	0	3	26
E-AGG/Hm5	3	36	4	13	2	58
E-AGG/Hm6	3	28	3	2	0	36
E-AGC/Hm1	0	8	15	1	1	25
E-AGC/Hm2	0	4	17	1	0	22
E-AGC/Hm3	2	30	5	1	1	39
E-AGC/Hm4	39	10	1	7	3	60
E-AGC/Hm5	18	10	5	2	20	55
E-AGC/Hm6	14	8	0	5	1	28

Table 4.7: Number of Methylated Sites Estimated by MS-AFLP Analysis

DNA Banding Description	Sum of Bands Scored	% of Total Analyzed
EcoRI/MspI-digested		
Same	576	26.4%
Unique	569	26.1%
Prominent	188	8.6%
EcoRI/MspI-digested		
Same	576	26.4%
Unique	151	6.9%
Prominent	124	5.7%
Total Scored Bands	**2184**	
Total Estimated Methylated Sites	**757**	**52.0%**
Total Estimated Unmethylated Sites	**700**	**48.0%**
Total Sites Analyzed	**1457**	

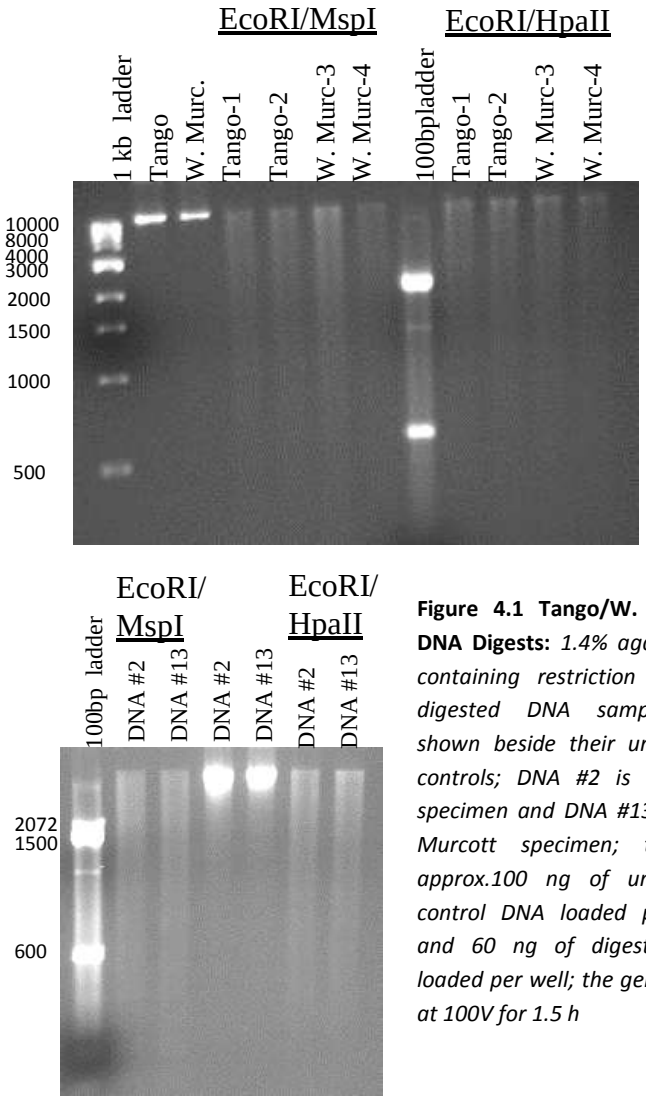


Figure 4.1 Tango/W. Murcott DNA Digests: *1.4% agarose gel containing restriction enzyme-digested DNA samples are shown beside their undigested controls; DNA #2 is a Tango specimen and DNA #13 is a W. Murcott specimen; there is approx.100 ng of undigested control DNA loaded per well, and 60 ng of digested DNA loaded per well; the gel was run at 100V for 1.5 h*

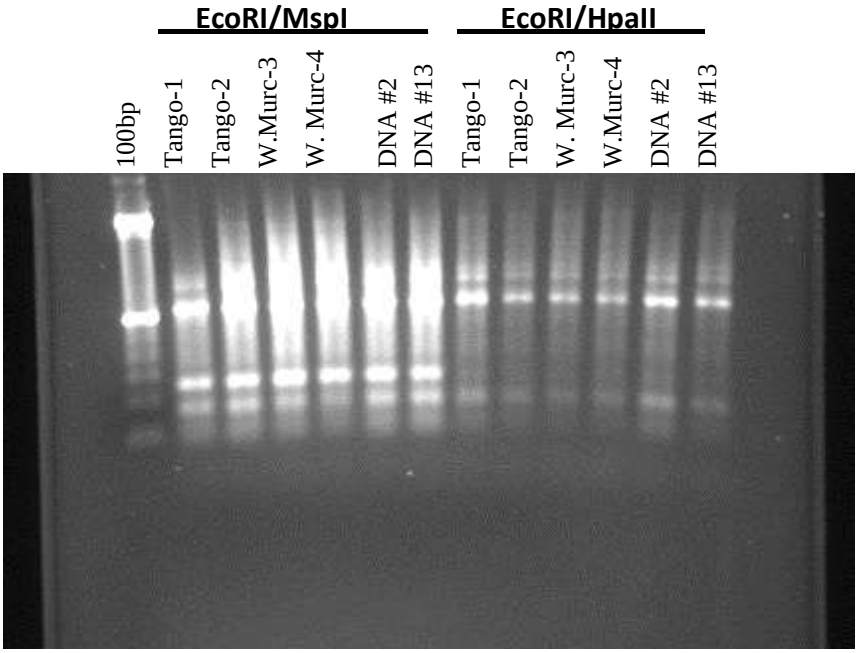


Figure 4.2 Preamplified PCR Products From DNA Digests: *1.4% agarose gel containing pre-selective amplified DNA digested with EcoRI/MspI or EcoRI/HpaII restriction enzyme combinations and ligated with adapters; note that banding patterns are alike between DNA samples digested with the same restriction enzymes*

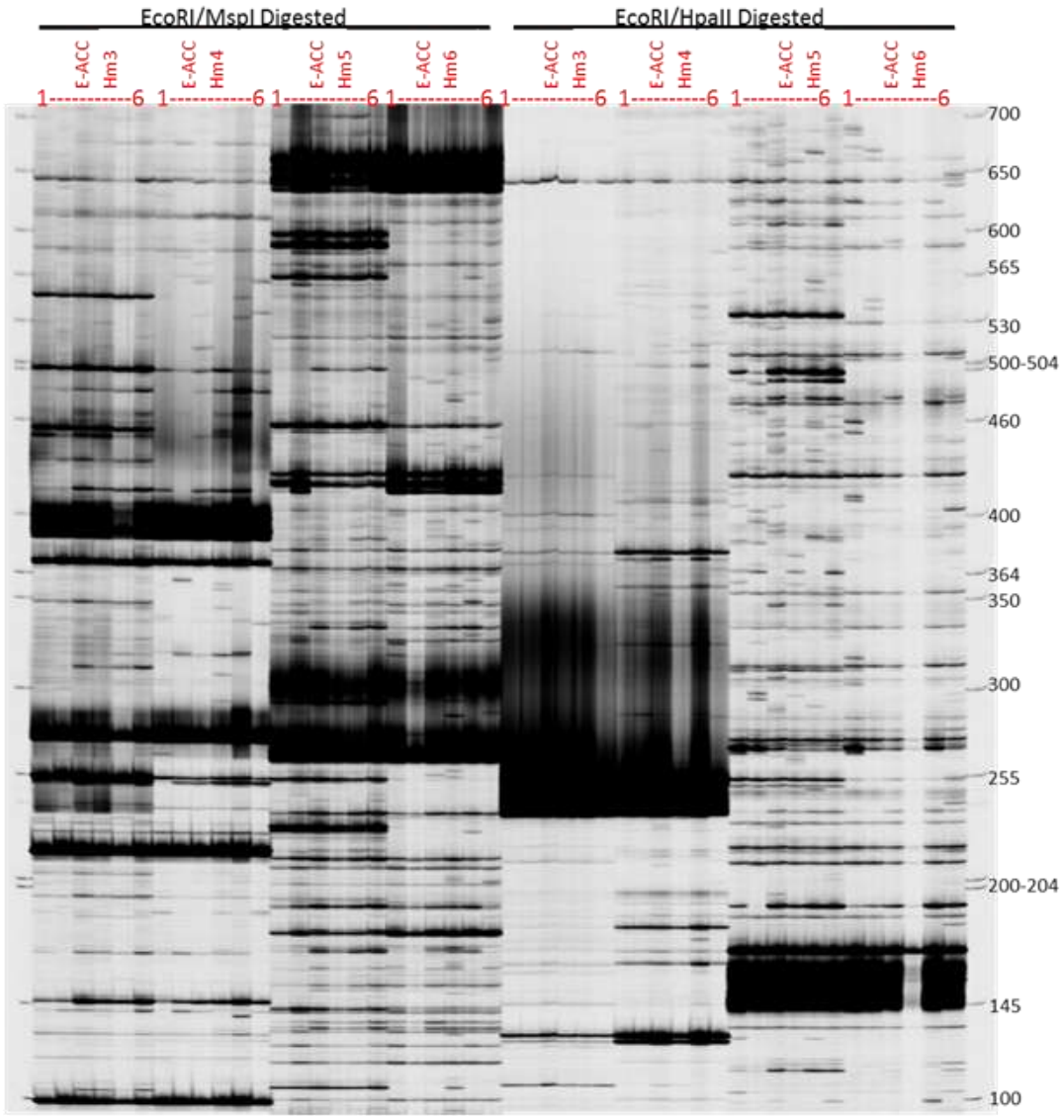


Figure 4.3 Methylation 5-700 Polyacrylamide Gel Image: *MS-AFLP analysis using 7% polyacrylamide gels; DNA samples are numbered 1-6, with each number representing each DNA sample as follows: 1=Tango-1; 2=Tango-2; 3= DNA #2; 4=W.Murcott-3; 5=W. Murcott-4; 6=DNA #13;700 bp DNA ladder was used to measure DNA fragments; five different gels were run, each denoted by a number 1-5, each gel generated separate images for the 700 or 800 infrared-labeled EcoRI primer used in amplification of DNA fragments; EcoRI and MspI/HpaII primerset combinations are denoted in red for each column (1 column contains 6 lanes); columns are divided between EcoRI/MspI and EcoRI/HpaII restriction enzyme digested DNA samples, with leftmost columns containing EcoRI/MspI DNA digests and rightmost columns containing EcoRI/HpaII DNA digests*

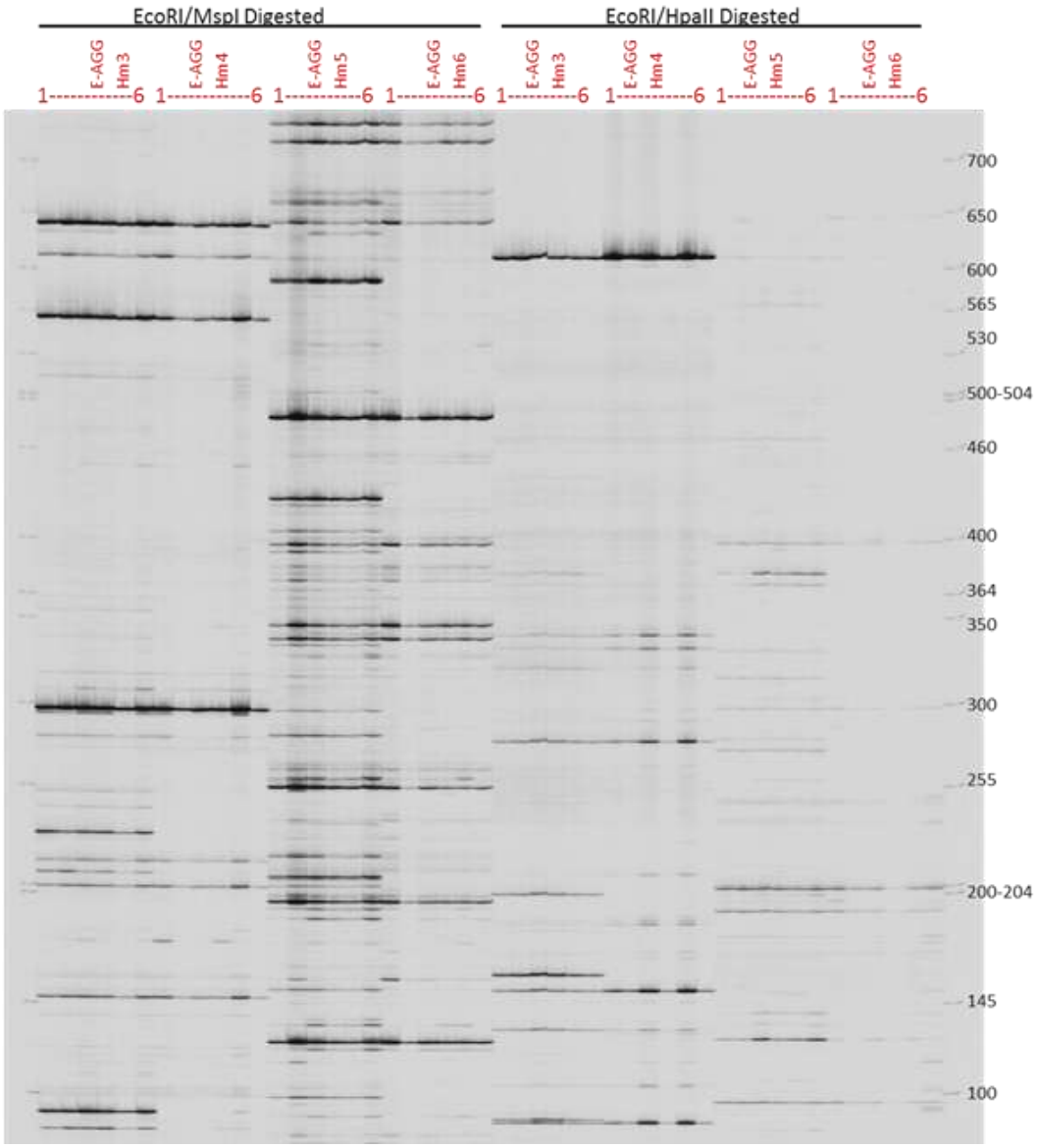


Figure 4.4 Methylation 5-800 Polyacrylamide Gel Image: *MS-AFLP analysis using 7% polyacrylamide gels; DNA samples are numbered 1-6, with each number representing each DNA sample as follows: 1=Tango-1; 2=Tango-2; 3= DNA #2; 4=W.Murcott-3; 5=W. Murcott-4; 6=DNA #13;700 bp DNA ladder was used to measure DNA fragments; gels were run at five different runs, each run denoted by a number 1-5, and viewed at 700 or 800, so that the 700 or 800 infrared-labeled EcoRI primer used in amplification of DNA fragments could be viewed; EcoRI and MspI/HpaII primerset combinations are denoted in red for each column (1 column contains 6 lanes); columns are divided between EcoRI/MspI and EcoRI/HpaII restriction enzyme digested DNA samples, with leftmost columns containing EcoRI/MspI DNA digests and rightmost columns containing EcoRI/HpaII DNA digests*

Conclusion to Dissertation

Key Difference Between W. Murcott and Tango

Experimental evidence suggests that, while W. Murcott and Tango do not differ significantly from one another in DNA markers, overall DNA methylation, or site specific methylation, they do differ from one another in their chromosomal configurations in the metaphase and anaphase stages of meiosis. The abnormal chromosomal configurations that have been documented in Tango gametic cells versus the normal chromosomal configurations that have been documented in W. Murcott gametic cells suggest that Tango, unlike W. Murcott, produces genetically anomalous gametes.

Male Sterility Linked to Seedlessness

As noted previously in chapter 2, male sterility in a plant can be caused by the production of anomalous gametes by the anthers. In the chromosome analysis, gametic cells were derived exclusively from anther tissue. Tango pollen has been shown to have low viability, which may be manifest as male sterility, and which could be a major contributor to its seedless phenotype. If the male gametic cell produced by Tango contains an irregular amount of genetic material, any normal female gamete that it may fuse with will result in an unviable zygote. This could explain not only the failure to produce seeds when Tango is self-pollinated, it could also explain why compatible varieties of *Citrus* cross pollinated with Tango fail to produce seeds. Though it is likely that male sterility is

a contributing factor to the seedless phenotype seen in Tango, it is unlikely to be the sole reason for this seedless characteristic.

Other Factors That May Contribute to Seedlessness

As noted previously in chapter 1, Tango does not set seed even when cross pollinated with compatible varieties of *Citrus.* While pollen from these varieties normally induces seed production in other compatible varieties, such as in W. Murcott, it fails to do so in Tango. Since there is no reason to believe that there is anything genetically abnormal in the pollen grains derived from the other *Citrus* varieties, it is logical to assume that the production of viable zygotes is inhibited by additional mechanisms. One possible factor could be female sterility related to genetically abnormal oocytes produced by Tango. If a female gamete containing irregular genetic material were to fuse with a male gamete containing normal genetic material, the resulting zygote would be genetically unbalanced, leading to its subsequent abortion. Failure to produce viable zygotes would result in seedlessness.

Further Chromosomal Analysis

Similar chromosomal analysis comparisons could be made on gametic cells derived from the female tissues of W. Murcott and Tango, using the same chromosome stain technique that was used on the male gametes. If meiotic chromosomal configurations in the metaphase and anaphase stages are abnormal in Tango gametic cells, yet normal in W. Murcott gametic cells, this would provide strong evidence for female sterility in Tango. Further analysis could also identify which chromosome/s may be involved in a translocation, as indicated by quadrivalent formations seen in Tango chromosomal

spreads. In order to target individual chromosomes unique markers, such as BAC clones that have a low content of repeated sequences, would need to be discovered on each chromosome that could be used as probes for fluorescent *in situ* hybridization (FISH). As technology improves with time, confocal microscopy may produce better quality images that could help in identifying chromosomes; as of now, the resolution is inadequate.

Overview of CIRE1 Retrotransposon Activity

No evidence was found to suggest that the CIRE1 retrotransposon was activated and replicated in response to stress induced by gamma radiation exposure, based on the analyses that were conducted to measure copy number of CIRE1 in the Tango genome when compared to the W. Murcott genome. If the CIRE1 retrotransposon had replicated and transposed in response to gamma radiation, the copy number of the CIRE1 retrotransposon would be higher in Tango than in W. Murcott. Tests using qPCR, a highly sensitive tool used to detect copy numbers of a DNA sequence, showed that the copy numbers of the targeted right-LTR region of CIRE1 were identical between W. Murcott and Tango. It is therefore unlikely that the CIRE1 retrotransposon replicated and transposed within the mandarin genome in response to the stress conditions brought on by gamma radiation exposure. Moreover, the unlikelihood that there was activation of the CIRE1 retrotransposon suggests that gene mutation due to CIRE1 insertion is also unlikely. Thus the seedless phenotype seen in Tango is probably not due to a CIRE1 insertion that alters one or more genes that are crucial to the fertility of the tree.

Estimate of CIRE1 Retrotransposon Copy Number

Interestingly, data from qPCR, as well as two separate analyses used to support its findings, all indicated that the CIRE1 retrotransposon exists in a much lower copy number within the *Citrus* genome than stated by Rico-Cabanas *et al* (2006). The technique used by Rico-Cabanas *et al* for absolute quantification of the CIRE1 retrotransposon within the *C. sinensis* genome was slot blot hybridization, a crude method that is more appropriate for semi-quantitative estimates of copy number. Unlike slot blot hybridization, qPCR is a highly sensitive, accurate, and reliable method for determining absolute copy numbers of a DNA sequence within a genome. Rico-Cabanas *et al* stated a CIRE1 copy number of 2,200 within the *C. sinensis* genome, while qPCR analysis estimated no more than 14 copies existing within the *C. sinensis* genome. This staggering difference in copy number estimation was later supported by analyses using Southern blot and BLAST, both of which indicated that CIRE1 exists in low copy number within the evaluated *Citrus* genomes. Based on the results from qPCR, Southern blot, and BLAST against reference *Citrus* genomes analyses, it is clear that the number of CIRE1 retrotransposon copies within the *C. sinensis* genome was overestimated by Rico-Cabanas et al.

Further Retroelement Analysis

Though our evaluations of the CIRE1 retrotransposon do not indicate that it was activated in response to stress induced by exposure to gamma radiation, this does not rule out the possibility of activation of other, as yet unidentified retrotransposons. As the availability and expansion of large insert libraries and breakthroughs in genome sequencing are made towards uncovering what is yet unknown in the genetics of *Citrus*, this will allow for the

discovery of new potentially active retroelements. Further experimentation as to whether or not other retrotransposons were activated and generated copies of themselves within the Tango genome in response to gamma radiation stress would be required to determine if this is indeed the case. Such data could be obtained by sequencing, assembly and comparison of the Tango and W. Murcott genomes.

Overview of Methylated Cytosine Findings

No evidence was found to suggest that there are significant differences in the percentage of methylated cytosine between W. Murcott and Tango genomes, according to data obtained from LUMA analysis. Moreover, a separate technique used to target specific methylated cytosine sites within the genome, known as methyl-sensitive AFLP, did not show differences in methylation among 1457 target sites analyzed between W. Murcott and Tango. This is only a small fraction of all methylated sites in the genome, so changes in methylation at specific sites remain possible, but it is clear that these altered sites comprise no more than a small portion of the genome. It is therefore unlikely that the seedless phenotype seen in Tango is attributed to an alteration in methylated cytosine patterns within its genome.

LUMA Pyrosequencing and Methyl-Sensitive AFLP Conclusions

Though no significant differences in methylated cytosine percentages or sites were found between the genomes of W. Murcott and Tango, both analyses agreed that the percentage of methylated cytosines within leaf tissues of both *Citrus* varieties was around 52%. This agreement indicates that both tests were likely to have been performed accurately. In the LUMA pyrosequencing analysis, gynoecium had a lower percentage of methylated

cytosines than leaf tissue in both cultivars, but the differences were not statistically significant. When data on Tango and W. Murcott were combined the overall average percentage of methylated cytosines in W. Murcott and Tango gynoecium tissue was significantly lower than in young leaf tissue. This may indicate that more of the chromatin in gynoecium tissue is available for transcription. This is the first estimate of the level of methylation in the citrus genome.

Further Epigenetic Analysis

Although no significant differences in methylated cytosine patterns or in overall percentage of methylated cytosines within the genome could be found between W. Murcott and Tango, it is possible that other epigenetic alterations may have occurred within Tango in response to stress-inducing gamma radiation treatment. Other epigenetic alteration possibilities could be targeted in future studies, such as histone modification. It is possible that modifications to histones in response to stress induced by gamma radiation may have played a role in activating previously silenced genes, thus causing a seedless phenotype in Tango. Further experimentation could also be conducted to identify the specific location of the unmethylated histones; it would be of great interest to find out whether or not these histones are located in promoter regions and, if so, which genes are linked to these promoters.

This research has shown that seedlessness in Tango could be explained by chromosomal aberrations, however, there was no evidence to indicate that the CIRE1 retrotransposon was activated in response to stress caused by exposure to gamma rays, nor that there was

any significant epigenetic difference between W. Murcott and Tango regarding cytosine methylation.

Appendix A

A. tgtgaagata tgtgtatatg gacttaaatg aagaaactgg aaaagtagag agctgttgtg cagattgact tcagttgatg gggtctattt gatatcttcc gctgacgtgg catgaattcg ttataatctg ttataacgaa gcccagcaca ctgatataaa acagaggctg acgaatggtg cctcatactg tgctaatatc gtcttcattt gttcatcatt ttgcatagaa tctgtttcat tgtgatctgt gcataatctt ctgtaaatca cctgtgattg ggttagttct tgatcaataa taatgagtaa attttccttc taagtttgct ggtctgtgtc ttttgttttg tctctggttt agcttctttg ttctggtttt tcttgattcg attgtgttct tggtgtttct actcaccaag attcgtgagt cttgctgtgg tttaaactta aaatcttaac a

B. gatagaacc aagttatggc accttaggct tggtcatatg agtatcaagg ggctgcaaga gcttagtaaa caagggctac ttggtggg**ga tagaatccag caactg**gagt tttgtgaaaa ctgtattttt ggcaaatcac acaggtccaa gtttaacaaa ggggagcata tgtctaaaca agttttggac tatgcacata ttgaccattg gggacctgct caggtccctt ccttgagtgg tggcagatat ttcatgtcac ttattgatga ttactcaagg aaagtatgga tctacatact caagatcaaa gatcaagctt tggaaaaatt caaggtctgg aaatcattgg tagaaaatca gtctgacttc aagctgaaat gcttgaggac tgataatggc ttggaatttt gtagcaaagt atttgaagaa tattgccaaa aacatggcat aaagaggcac aaaacagtga ggttcacccc ccagcagaat ggcttggctg agaggatgaa caggaccttg gtggataaaa ccagatg**tat gctgataaat tcaaagct**tc caagaagttt ttgggcagaa gctgtcaata cagcaagtta tttagttaac agaagcccat cttctgcaat tggcttcaaa acaccagaag aattgtggaa tggaaaacca gcaaactacc aaaatctcag agtttttggg tgccctgctt acctccacat caatcaaggc aagctggaag caagggcctt gaaaggagtc tttgttggat at

C. aagtggct tgaggcaatg caagatgaaa tggagtcact tcaaagaaat cagacctgga cacttatacc aaatccagga aacaagaggt tgatcaactg taagtggatt ttcaagagga atgaaggaat tccagatgtt gaacctccca aatataaagc aaggctagtg gctaggggtt ttacacagag g**gaaggtgta gatttcaatg** aaatttttc tcctgtagtg aagcacagtt caataagaat cttattagcc atggtagctt tacttgactt ggagctggag cagatggatg taaaaacagc cttccttcat ggtaatcttg aagaacaaat attgatggca cagcctgaag gatttgagtg taaggataag gaggattatg tgtgtctgct gcataagtcc ttgtatggac tcaagcaatc ccctcgacag tggtacagaa ggtttgatga ttttatggtc tcaaagggct atcgcagaag cagatatgac agctgtgttt actttggtgg atcagatcaa ggtggagttg cttatttgct tctgtatgtt gacgatatgt taat**tgcaag caaatacaag tcag**agatag agaggttaaa gaatctgtta aaggctgagt ttgaaatgaa ggatcttggc aatgccaaaa gaatccttgg gatggacatt atcagagata gatctgctgg cactttattc ttgagtcaag ggaaatatat c

Appendix B: *Shown are the full-length sequences that were used in the BLAST analysis, including the full-length right LTR region of CIRE1 (A.), the full-lenth integrase gene found in CIRE1 (B.), and the full-length reverse transcriptase gene found in CIRE1 (C.). The full-length right LTR CIRE1 sequence can be found in the EMBL database under the accession number AM040263, while both the full-length integrase and the full-length reverse transcriptase genes were found by comparing amino acid sequences between CIRE1 and Tal-3 copia transposable element found in Arabidopsis thaliana; amino acid sequence matches in the regions corresponding to the genes of interest were then converted into nucleotide sequences*

Appendix B

Calculations to Determine Copy Number of CIRE1 Retrotrsansposon in qPCR

The size of the *Citrus* genome is **3.82X10^(8) bp** equivalent to about **3.82X10^(-4) pg**

Concentration of *Citrus* DNA samples as measured by Nanodrop Spectrophotometer:

Parent Washington Navel: **80.99 ng/µl** Hart's Tardiff Valencia: **65.29 ng/µl** W. Murcott: **24.93 ng/µl**

Tango: **31.7 ng/µl** Abhayapuri Lime: **76.53 ng/µl**

Citrus DNA concentrations have been diluted by 1000-fold, then further diluted into 1:2, 1:4, 1:8, and 1:16 dilutions.

CIRE1 right-LTR DNA fragment is **247 bp** long; the plasmid it is cloned into is **3956 bp** long, meaning that the insert alone makes up **5.88%** of the insert + plasmid construct. The concentration of undiluted plasmid + insert used in making the standard curve as measured by Nanodrop Spectrophotometer: **19.6 ng/µl**; dilutions of the original stock concentration used in making the standard were performed ten-fold. The amount of DNA insert per dilution used in the qPCR reaction was calculated as follows: **19.6 ng/µl** plasmid + insert **X 2 µl** used in reaction **X 5.88%** which is the percentage of the construct that is the DNA insert **X 10^(3)** to convert to picograms **= 2305 pg/µl** undiluted CIRE1 right-LTR DNA fragment used in the standard curve. Dilutions made for the curve affected the calculated concentration as follows:

dilution	concentration of insert
10^(-6)	**2.305 X 10^(-3) pg**
10^(-6) 1:2	**1.1525 X 10^(-3) pg**
10^(-6) 1:4	**5.762X 10^(-4) pg**
10^(-6) 1:8	**2.8812 X 10^(-4) pg**
10^(-6) 1:16	**1.4406 X 10^(-4) pg**
10^(-6) 1:32	**7.203 X 10^(-5) pg**
10^(-6) 1:64	**3.6015 X 10^(-5) pg**
10^(-8)	**2.305 X 10^(-5) pg**

Based on normalization CT values, the most comparable dilutions to use in calculating copy number for each *Citrus* variety were as follows:

Parent Washington Navel: 1:4 dilution; **2.02475 X 10^(-2) ng/µl**

Hart's Tardiff Valencia: 1:4 dilution; **1.63225 X 10^(-2) ng/µl**

W. Murcott: 1:2 dilution; **1.2465 X 10^(-2) ng/µl**

Tango: 1:2 dilution; **1.585 X 10^(-2) ng/µl**

Abhayapuri Lime: 1:8 dilution; **9.56625 X 10^(-3) ng/µl**

Calculations to Determine Copy Number of CIRE1 Retrotransposon in qPCR

Parent Washington Navel: **20.2475 pg/µl** genome in 1:4 dilution; **X 2** µl used in reaction **= 40.5 pg genome in 1:4 dilution**. Based on corresponding CT value, there is **3.226 X 10^(-4) pg LTR fragment** present in reaction.

Thus to calculate copy number from these values:

Convert picograms of LTR fragment present in the reaction into grams of LTR fragment present in the reaction:
(**3.226 X 10^(-4) pg**)(1g/1 X 10^(12)pg) = **3.226 X 10^(-16) g** of LTR.

Convert this value into the number of base pairs of the LTR fragment present in the reaction based on mass (in grams) of a base pair: (**3.226 X 10^(-16) g**)(1 bp/1.096 X 10^(-21) g) = **294343.0657 bp** of LTR.

Divide the total number of base pairs of LTR present in the reaction with the number of base pairs found in one copy of the LTR fragment to determine how many total LTR fragments are in the reaction:
(**294343.0657 bp**)/(**247 bp** length of LTR fragment) = **1191.67 copies LTR in reaction**.

Convert picograms of Parent Washington Navel genome present in the reaction into grams of Parent Washington Navel genome present in the reaction:
(**40.5 pg**)(1g/1 X 10^(12) pg) = **4.05 X 10^(-11) g** of genome.

Convert this value into the number of base pairs of Parent Washington Navel present in the reaction based on mass (in grams) of a base pair:
(**4.05 X 10^(-11) g**)(1 bp/1.096 X 10^(-21)g) = **3.6953 X 10^(10) bp** of genome.

Divide the total number of base pairs of Parent Washington Navel genome present in the reaction with the number of base pairs found in one copy of the Parent Washington Navel genome to determine how many total copies of genome are in the reaction: (**3.6953 X 10^(10) bp**)/(**3.82X10^(8) bp** length of *Citrus* genome) = **96.73 copies of genome in reaction.**

Divide the total number of LTR copies present in the reaction with the total number of copies of Parent Washington navel genome in the reaction to find the average number of copies of LTR fragment present per copy of genome: (**1191.67 copies LTR**)/(**96.73 copies genome**) ≈ **12 copies of LTR fragment region per genome of Parent Washington Navel**

Calculations to Determine Copy Number of CIRE1 Retrotransposon in qPCR

Hart's Tardiff Valencia: **16.3225 pg/µl** genome in 1:4 dilution; **X 2** µl used in reaction = **32.65 pg genome in 1:4 dilution**. Based on corresponding CT value, there is **1.783 X 10^(-4) pg LTR fragment** present in reaction. Thus to calculate copy number from these values: (**1.783 X 10^(-4) pg**)(1g/1 X 10^(12)pg) = **1.783 X 10^(-16) g** of LTR. (**1.783 X 10^(-16) g**)(1 bp/1.096 X 10^(-21) g) = **162682.4818 bp** of LTR. (**162682.4818 bp**)/(**247 bp** length of LTR fragment) = **659 copies LTR in reaction**. (**32.65 pg**)(1g/1 X 10^(12) pg) = **3.27 X 10^(-11) g** of genome. (**3.27 X 10^(-11) g**)(1 bp/1.096 X 10^(-21)g) = **2.98 X 10^(10) bp** of genome. (**2.98 X 10^(10) bp**)/(**3.82X10^(8) bp** length of *Citrus* genome) = **78 copies of genome in reaction**. (**659 copies LTR**)/(**78 copies genome**) ≈ **8 copies of LTR fragment region per genome of Hart's Tardiff Valencia**

W. Murcott: **12.465 pg/µl** genome in 1:2 dilution; **X 2** µl used in reaction = **24.93 pg genome in 1:2 dilution**. Based on corresponding CT value, there is **2.263 X 10^(-4) pg LTR fragment** present in reaction. Thus to calculate copy number from these values: (**2.263 X 10^(-4) pg**)(1g/1 X 10^(12)pg) = **2.263 X 10^(-16) g** of LTR. (**2.263 X 10^(-16) g**)(1 bp/1.096 X 10^(-21) g) = **206478.1022 bp** of LTR. (**206478.1022 bp**)/(**247 bp** length of LTR fragment) = **836 copies LTR in reaction**. (**24.93 pg**)(1g/1 X 10^(12) pg) = **2.493 X 10^(-11) g** of genome. (**2.493 X 10^(-11) g**)(1 bp/1.096 X 10^(-21)g) = **2.275 X 10^(10) bp** of genome. (**2.275 X 10^(10) bp**)/(**3.82X10^(8) bp** length of *Citrus* genome) = **60 copies of genome in reaction**. (**836 copies LTR**)/(**60 copies genome**) ≈ **14 copies of LTR fragment region per genome of W. Murcott**

Tango: **15.85 pg/µl** genome in 1:2 dilution; **X 2** µl used in reaction = **31.7 pg genome in 1:2 dilution**. Based on corresponding CT value, there is **2.892 X 10^(-4) pg LTR fragment** present in reaction. Thus to calculate copy number from these values: (**2.892 X 10^(-4) pg**)(1g/1 X 10^(12)pg) = **2.892 X 10^(-16) g** of LTR. (**2.892 X 10^(-16) g**)(1 bp/1.096 X 10^(-21) g) = **263868.61 bp** of LTR. (**263868.61 bp**)/(**247 bp** length of LTR fragment) = **1068 copies LTR in reaction**. (**31.7 pg**)(1g/1 X 10^(12) pg) = **3.17 X 10^(-11) g** of genome. (**3.17 X 10^(-11) g**)(1 bp/1.096 X 10^(-21)g) = **2.9 X 10^(10) bp** of genome. (**2.9 X 10^(10) bp**)/(**3.82X10^(8) bp** length of *Citrus* genome) = **76 copies of genome in reaction**. (**1068 copies LTR**)/(**76 copies genome**) ≈ **14 copies of LTR fragment region per genome of Tango**

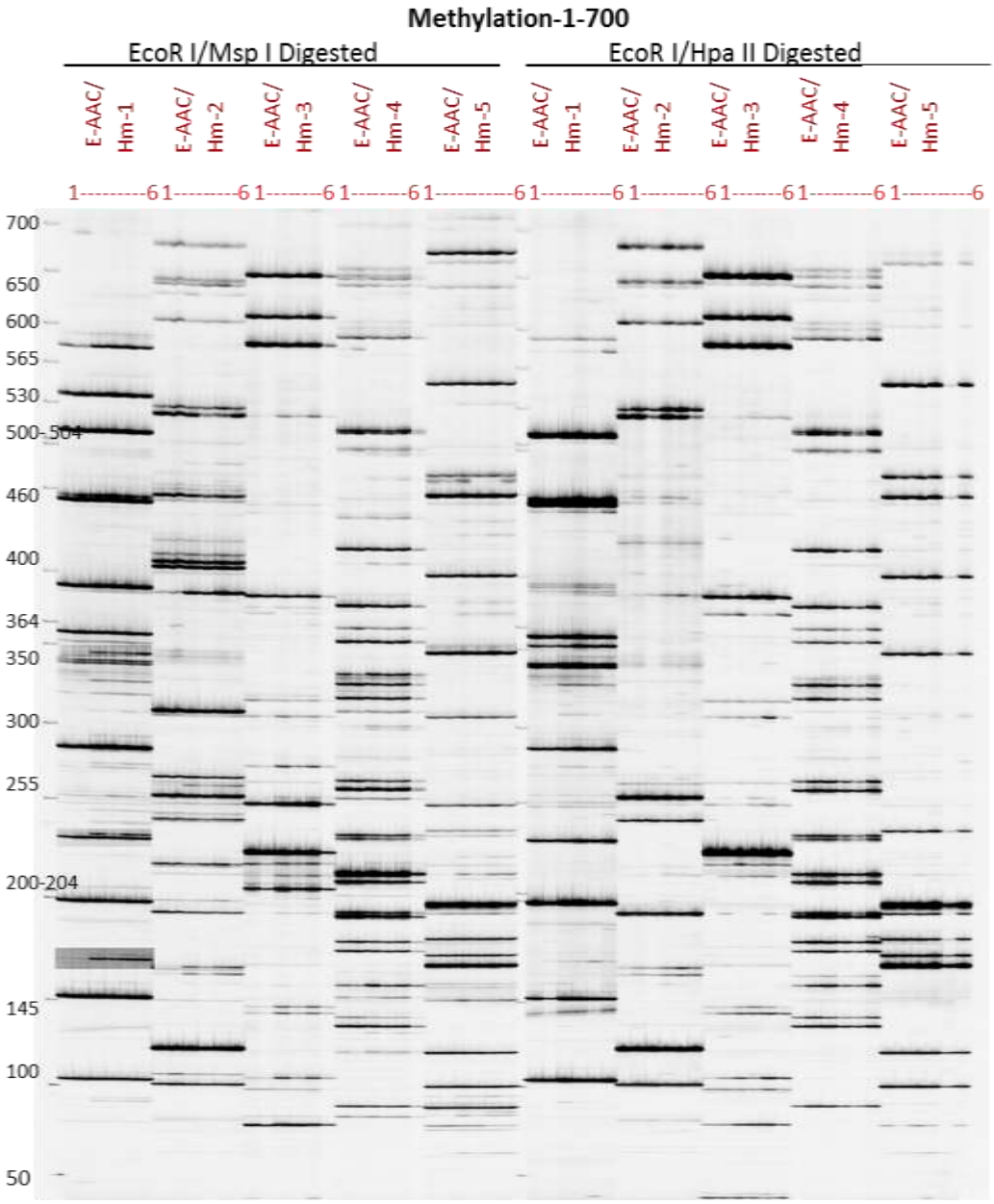
Abhayapuri Lime: **9.5663 pg/µl** genome in 1:8 dilution; **X 2** µl used in reaction = **19.13 pg genome in 1:8 dilution**. Based on corresponding CT value, there is **1.097 X 10^(-5) pg LTR fragment** present in reaction. Thus to calculate copy number from these values: (**1.097 X 10^(-5) pg**)(1g/1 X 10^(12)pg) = **1.097 X 10^(-17) g** of LTR. (**1.097 X 10^(-17) g**)(1 bp/1.096 X 10^(-21) g) = **10009.12 bp** of LTR. (**10009.12bp**)/(**247 bp** length of LTR fragment) = **41 copies LTR in reaction**. (**19.3 pg**)(1g/1 X 10^(12) pg) = **1.91 X 10^(-11) g** of genome. (**1.91 X 10^(-11) g**)(1 bp/1.096 X 10^(-21)g) = **1.75 X 10^(10) bp** of genome. (**1.75 X 10^(10) bp**)/(**3.82X10^(8) bp** length of *Citrus* genome) = **47 copies of genome in reaction**. (**41 copies LTR**)/(**47 copies genome**) ≈ **1 copy of LTR fragment region per genome of Abhayapuri Lime**

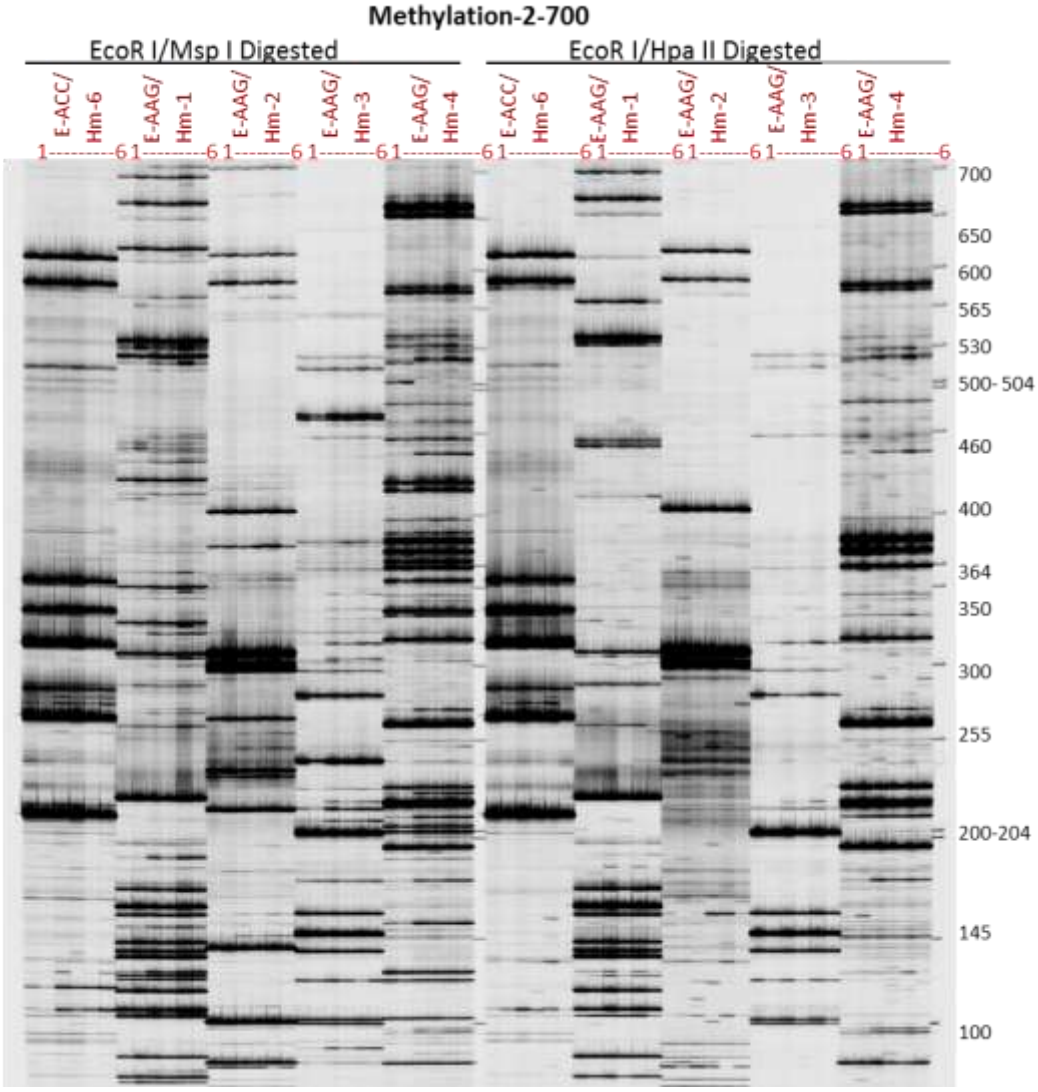
Sample ID	Duplicate	% Meth	Average	SD
Ctrl-Low	1	-4.3		
Ctrl-Low	2	-2.7	-3.5	± 1.14
Ctrl-Med	1	50.8		
Ctrl-Med	2	49.3	50.0	±1.02
Ctrl-High	1	98.9		
Ctrl-High	2	98.9	98.9	± 0.04
Tan-fb-15-2-10	1	51.7		
Tan-fb-15-2-10	2	50.6	51.2	± 0.74
Tan-yl-13-34-11	1	57.6		
Tan-yl-13-34-11	2	54.5	56.0	±2.23
WM-yl-1-20-13	1	58.5		
WM-yl-1-20-13	2	57.9	58.2	±0.38
WM-fb-1-20-13	1	57.8		
WM-fb-1-20-13	2	56.9	57.3	±0.65
WM-fb-13-14-14	1	54.0		
WM-fb-13-14-14	2	53.2	53.6	±0.52
Tan-fb-1-28-3	1	54.6		
Tan-fb-1-28-3	2	54.3	54.4	±0.23
Tan-yl-1-28-3	1	54.4		
Tan-yl-1-28-3	2	53.4	53.9	±0.65
WM-fb-12-48-3	1	50.1		
WM-fb-12-48-3	2	48.3	49.2	±1.29
WM-yl-13-14-14	1	57.6		
WM-yl-13-14-14	2	58.0	57.8	±0.27
Tan-fb-13-34-11	1	54.2		
Tan-fb-13-34-11	2	53.4	53.8	±0.53
Tan-yl-15-2-10	1	57.4		
Tan-yl-15-2-10	2	58.3	57.9	±0.62
WM-yl-12-48-4	1	67.8		
WM-yl-12-48-4	2	59.7	63.7	±5.70

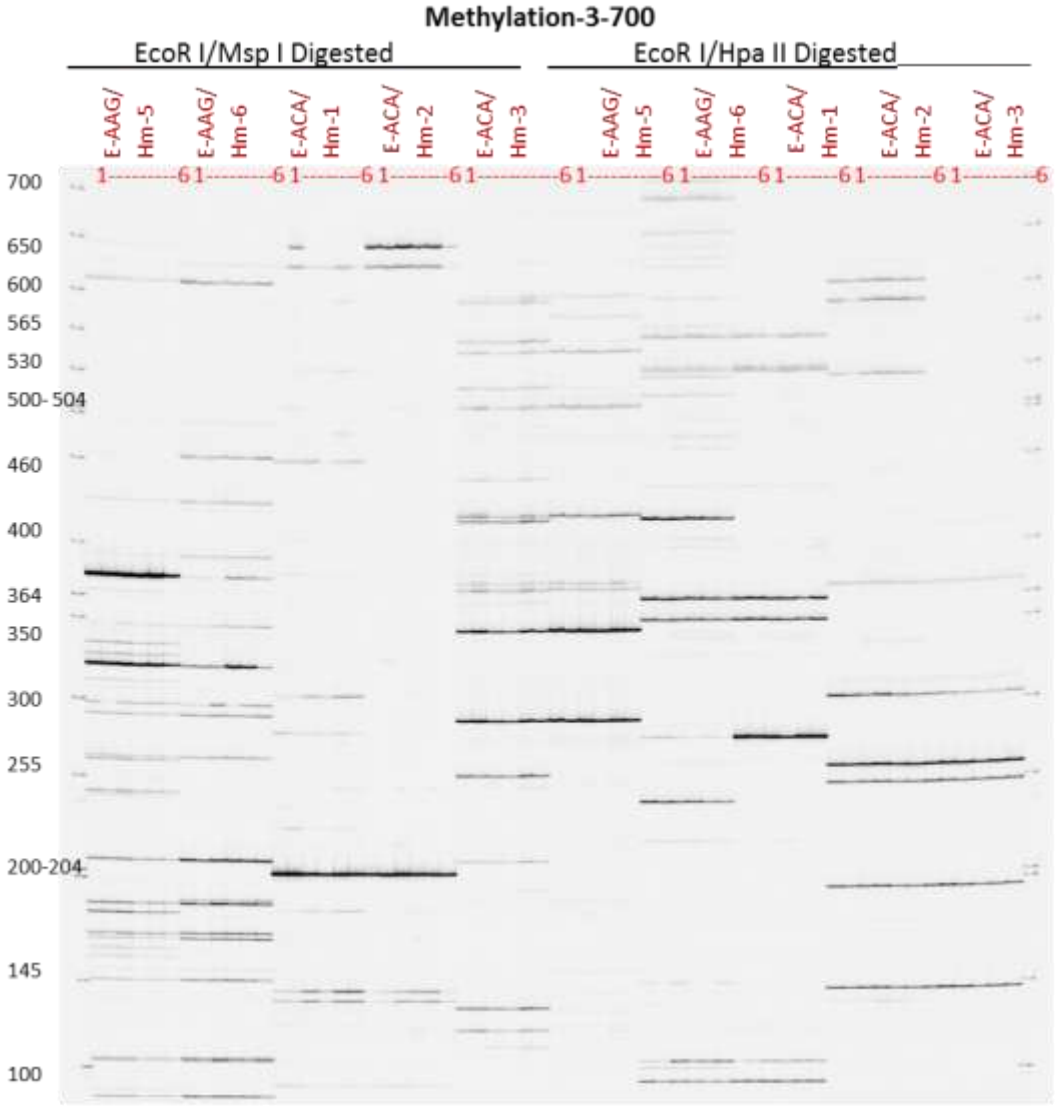
59.7

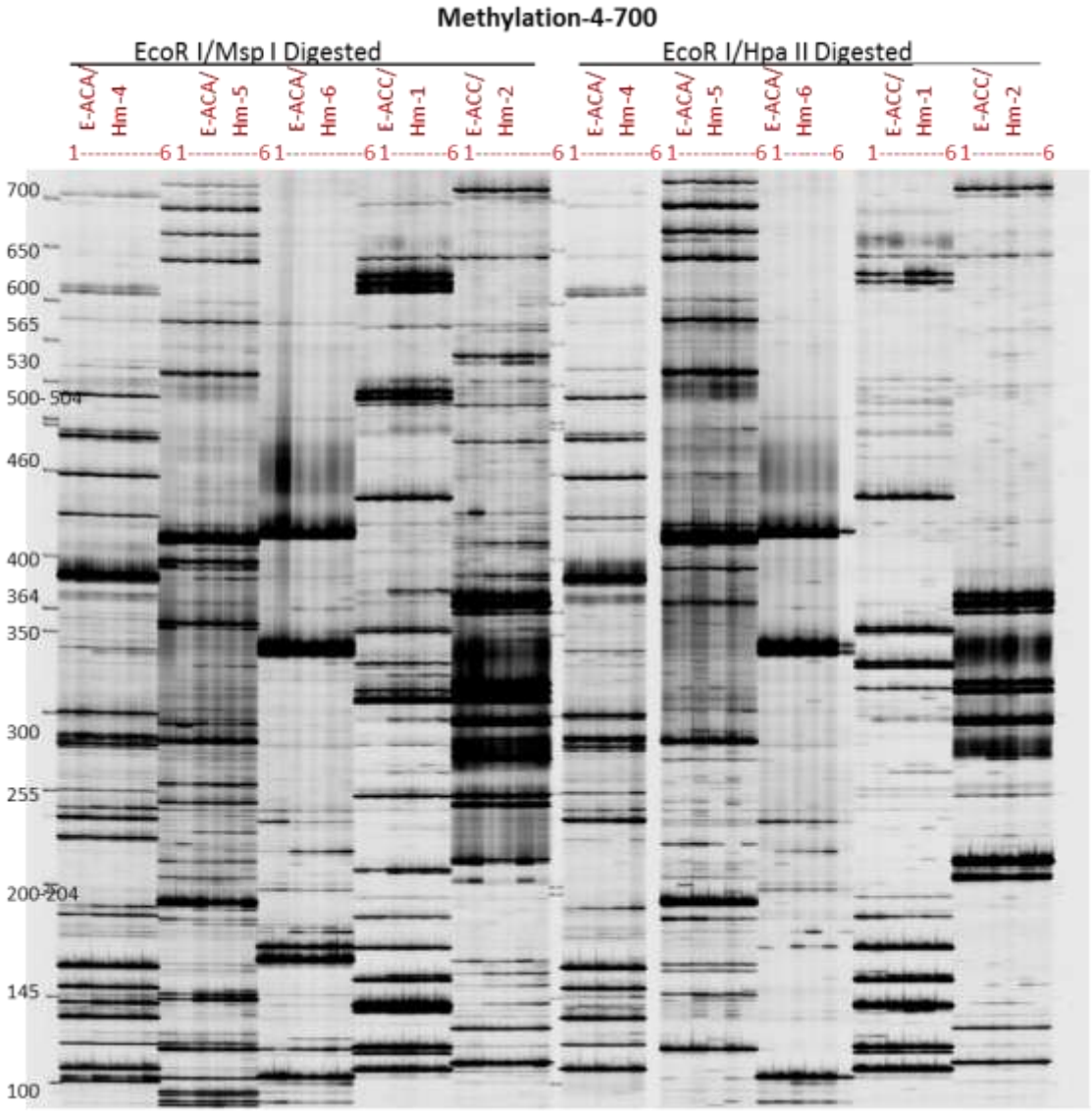
Appendix C: *LUMA analysis of methylated cytosine levels found in genomic DNA of multiple W. Murcott and Tango specimen samples; HpaII/MspI values for each sample tested is evaluated from the amount of signal from the unmethylated CCGG site compared to all the CCGG sites, or the percent unmethylated; the % methylation is calculated by subtracting this value from 1 and multiplying by 100, all samples are done in replicate to ensure accuracy of results, 3 controls are also included in the analysis to ensure that expected methylation percentage values are given in response to low, medium, and high methylated controls, respectively ; note the yellow highlighted outlier, it is suspiciously higher in % methylation value from its replicate; the methylation percentage average is shown as an average of the two replicates for each sample along with the standard deviation for each, the averages have been used in statistical analyses with the outlier thrown out and replaced with replicate 2 of the sample in question (highlighted in pink)*

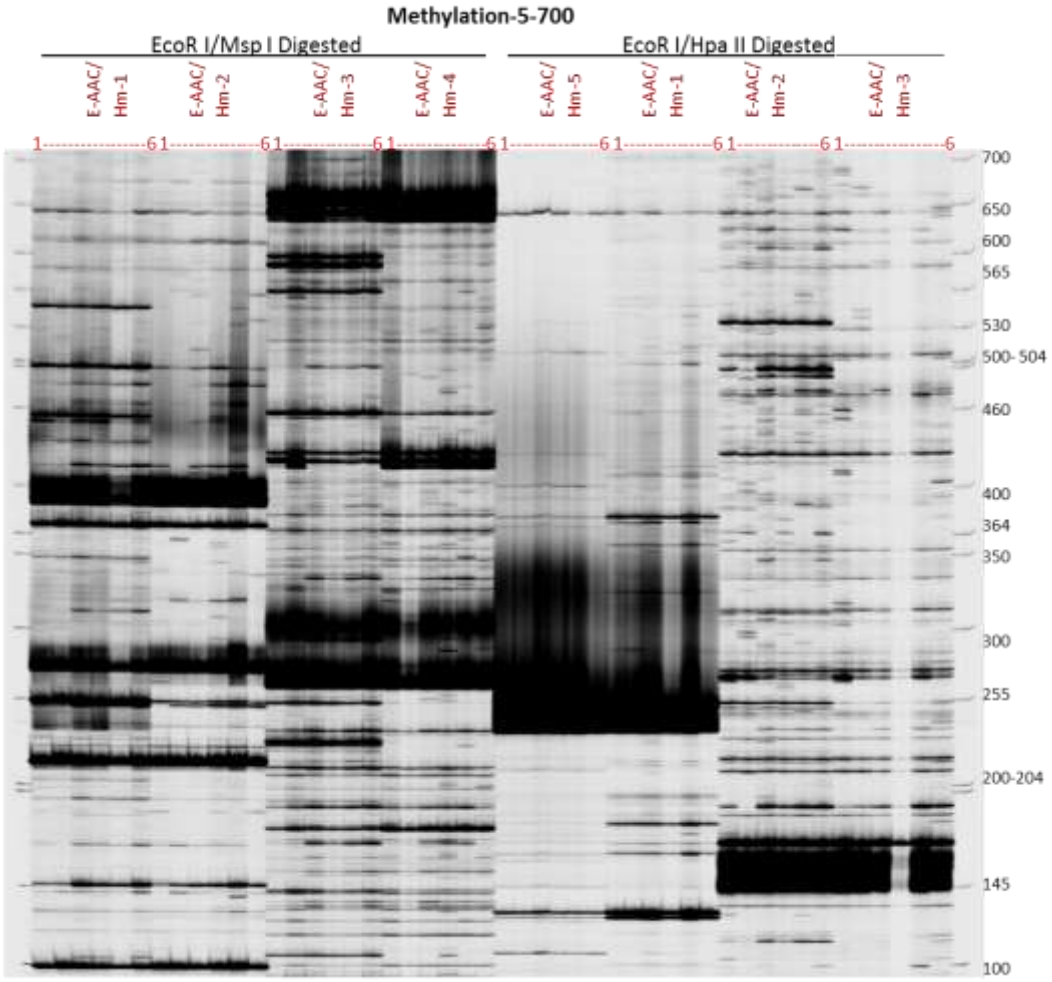
Appendix D

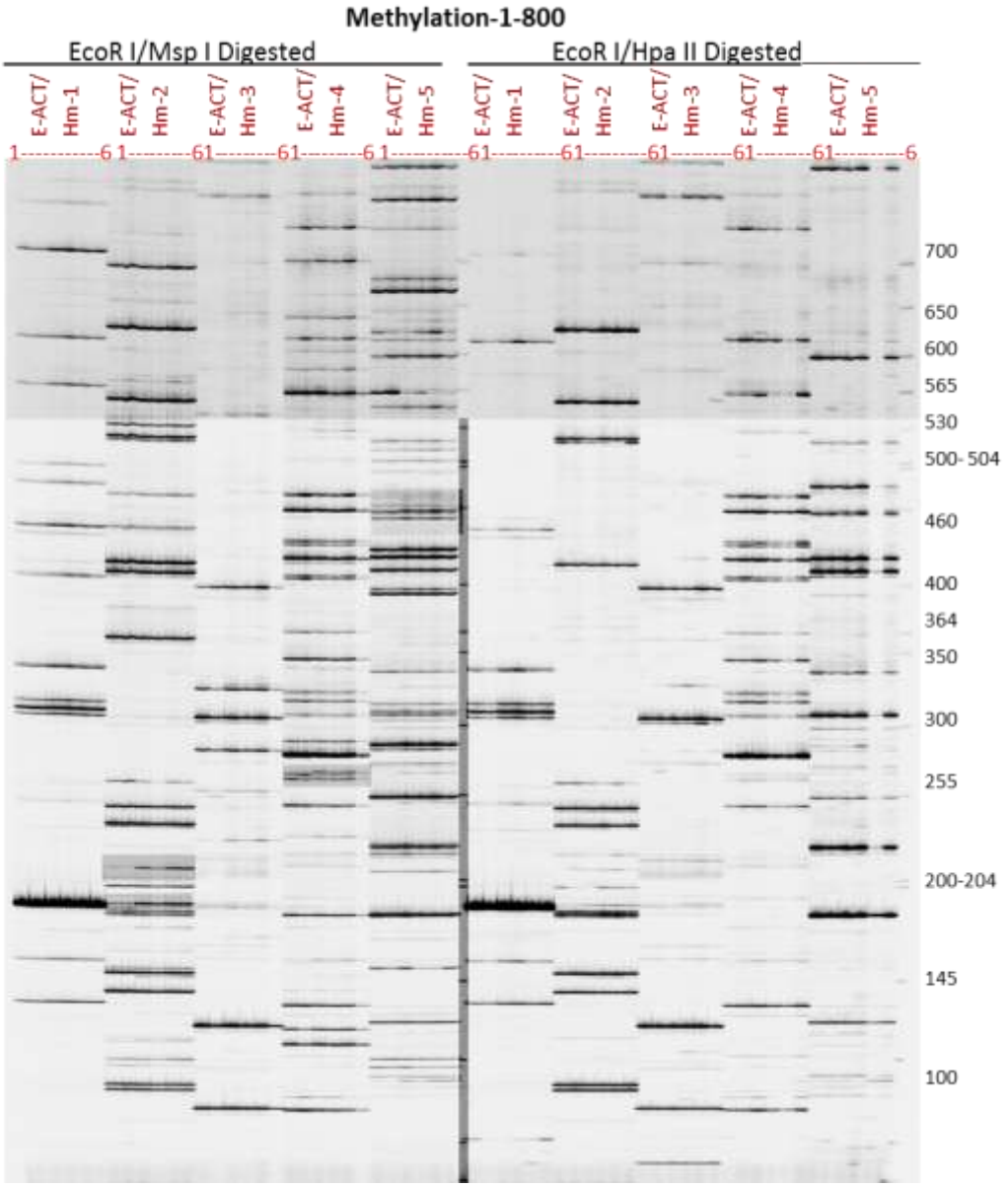

Methylation-1-700
EcoR I/Msp I Digested
EcoR I/Hpa II Digested
E-AAC/ Hm-1
E-AAC/ Hm-2
E-AAC/ Hm-3
E-AAC/ Hm-4
E-AAC/ Hm-5
1---------6 1---------6 1--------6 1--------6 1----------6 1---------6 1---------6 1--------6 1--------6 1---------6
700
650
600
565
530
500-504
460
400
364
350
300
255
200-204
145
100
50

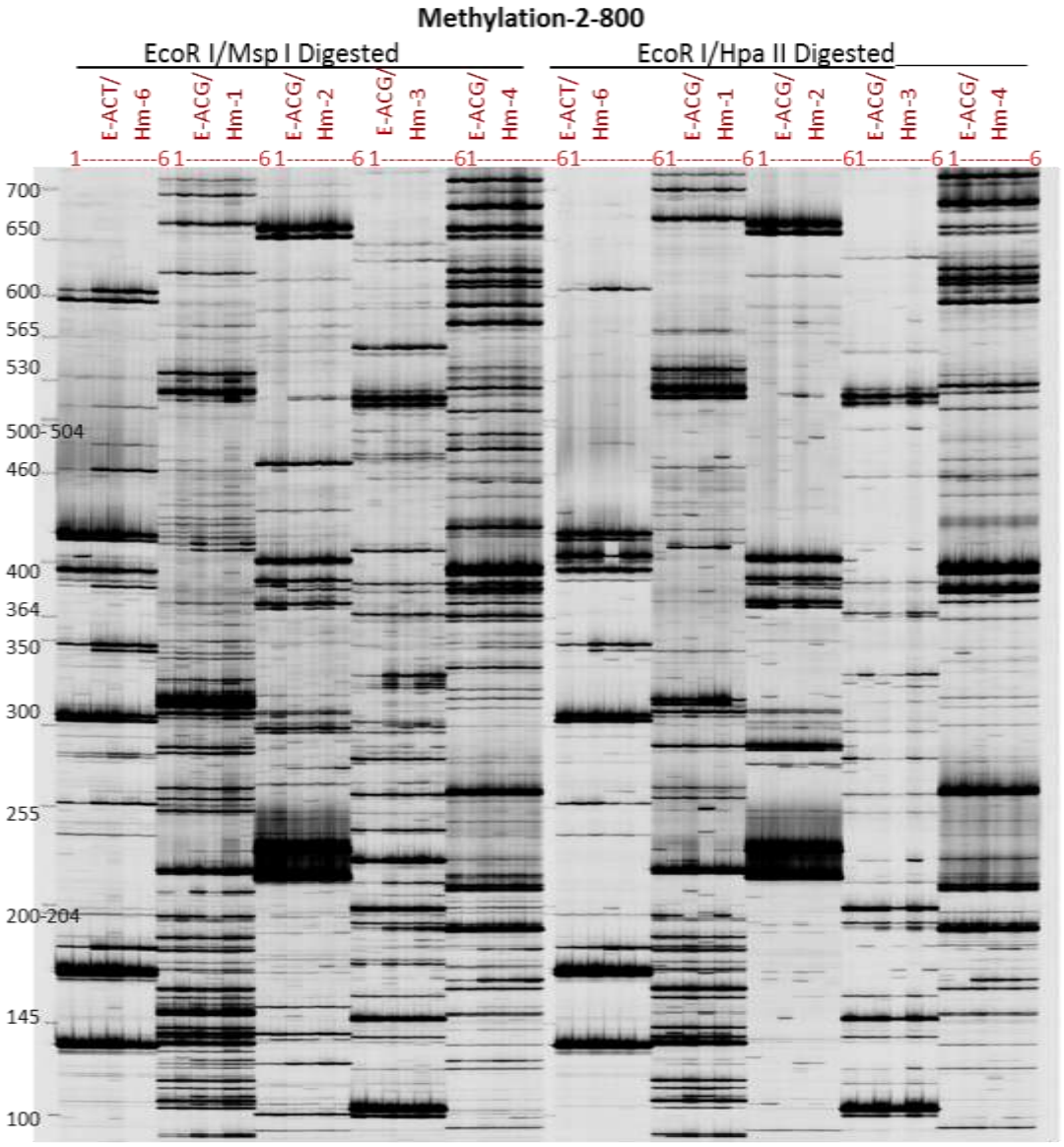

Methylation-2-700
EcoR I/Msp I Digested
EcoR I/Hpa II Digested
E-ACC/ Hm-6
E-AAG/ Hm-1
E-AAG/ Hm-2
E-AAG/ Hm-3
E-AAG/ Hm-4
1
6
700
650
600
565
530
500- 504
460
400
364
350
300
255
200-204
145
100

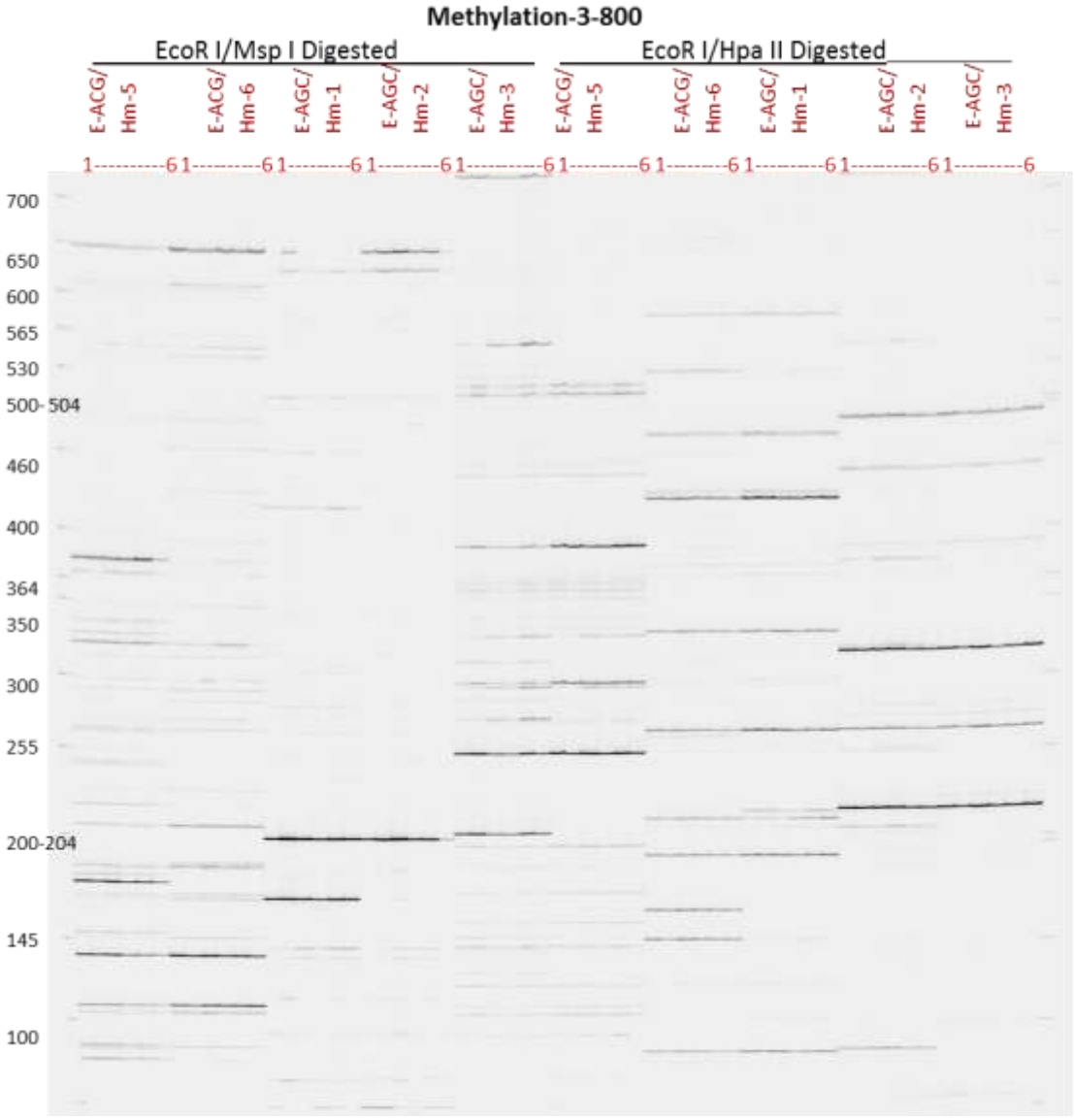

Methylation-3-700
EcoR I/Msp I Digested
EcoR I/Hpa II Digested
E-AAG/ Hm-5
E-AAG/ Hm-6
E-ACA/ Hm-1
E-ACA/ Hm-2
E-ACA/ Hm-3
E-AAG/ Hm-5
E-AAG/ Hm-6
E-ACA/ Hm-1
E-ACA/ Hm-2
E-ACA/ Hm-3
1 6 1 6 1 6 1 6 1 6 1 6 1 6 1 6 1 6 1 6
700
650
600
565
530
500- 504
460
400
364
350
300
255
200-204
145
100

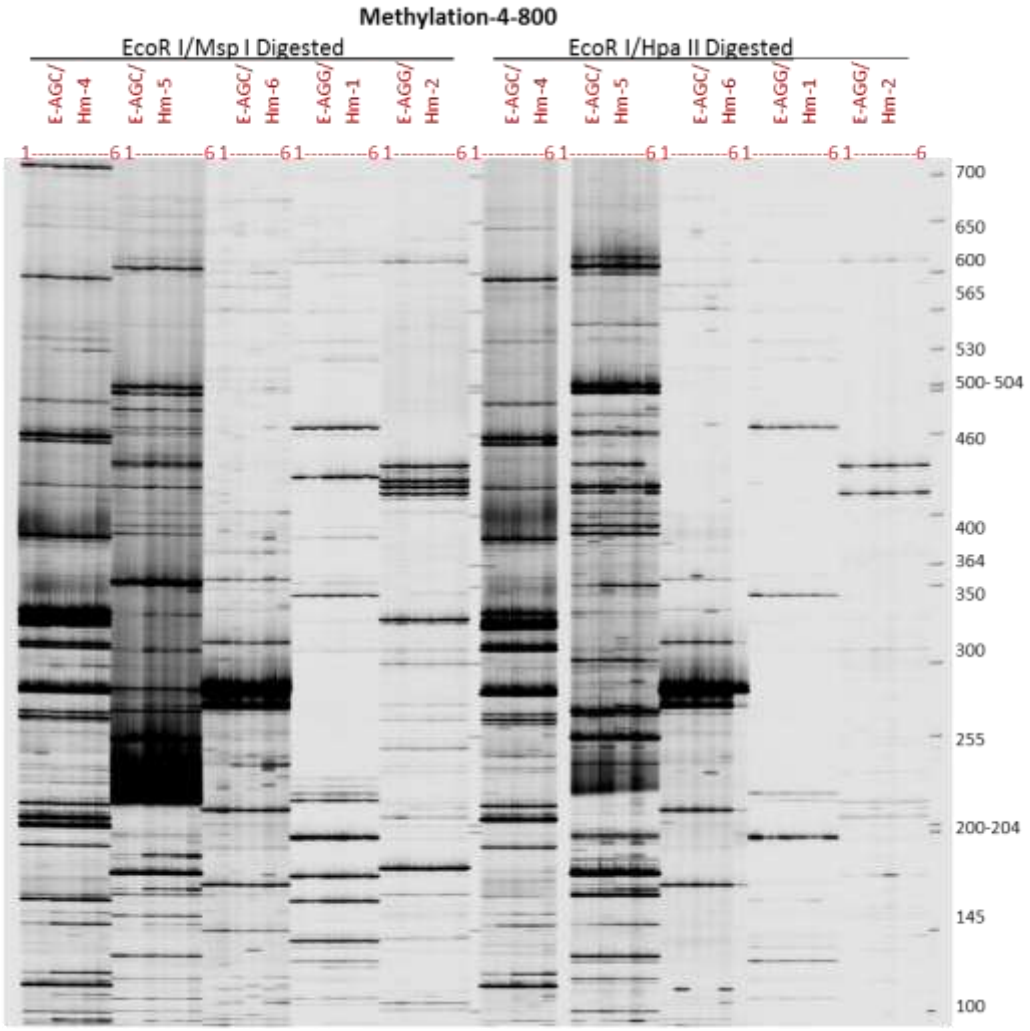

Methylation-4-700
EcoR I/Msp I Digested
EcoR I/Hpa II Digested
E-ACA/ Hm-4
E-ACA/ Hm-5
E-ACA/ Hm-6
E-ACC/ Hm-1
E-ACC/ Hm-2
E-ACA/ Hm-4
E-ACA/ Hm-5
E-ACA/ Hm-6
E-ACC/ Hm-1
E-ACC/ Hm-2
1----6
700
650
600
565
530
500- 504
460
400
364
350
300
255
200-204
145
100

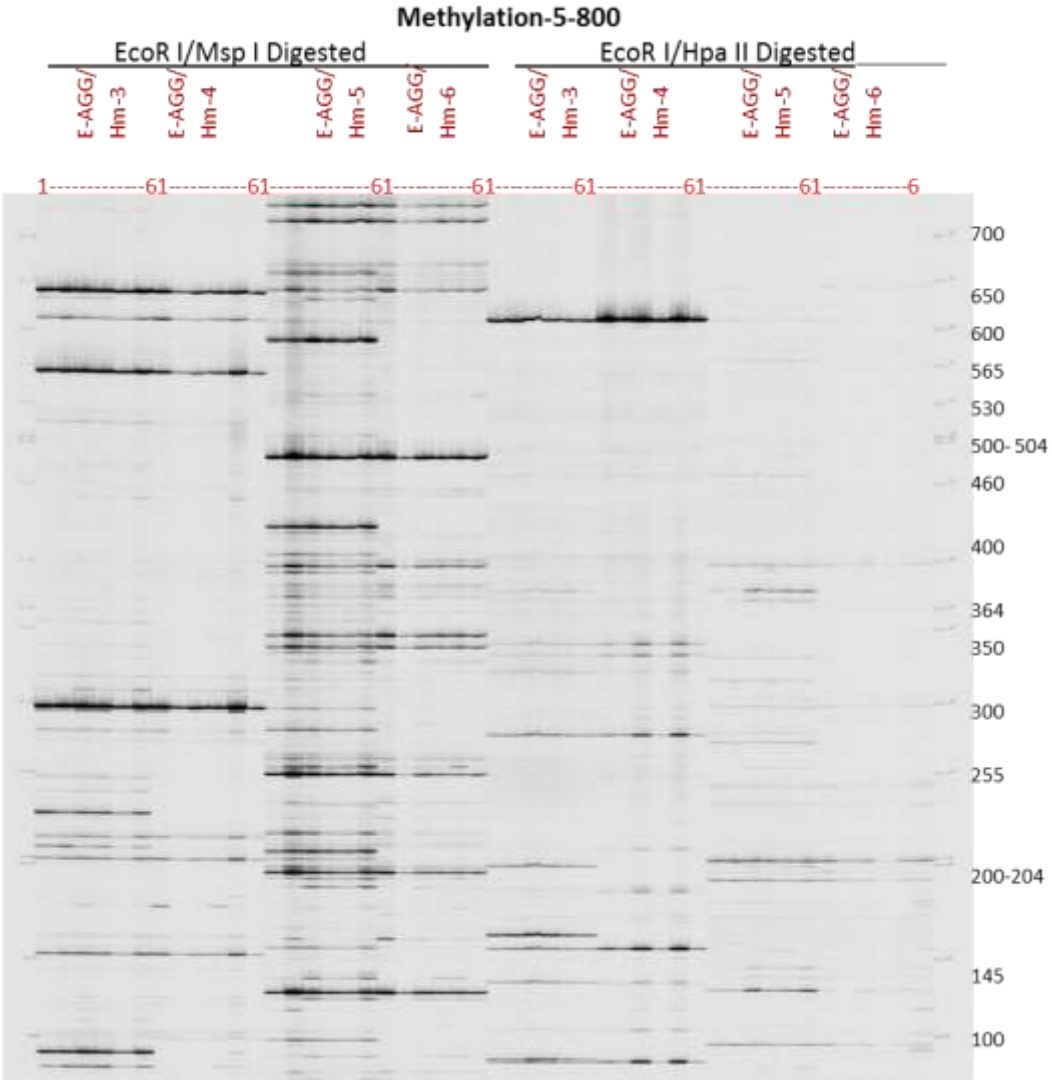

Methylation-5-700
EcoR I/Msp I Digested
EcoR I/Hpa II Digested
E-AAC/ Hm-1
E-AAC/ Hm-2
E-AAC/ Hm-3
E-AAC/ Hm-4
E-AAC/ Hm-5
E-AAC/ Hm-1
E-AAC/ Hm-2
E-AAC/ Hm-3
1 6 1 6 1 6 1 6 1 6 1 6 1 6 1 6
700
650
600
565
530
500- 504
460
400
364
350
300
255
200-204
145
100


Methylation-1-800
EcoR I/Msp I Digested
EcoR I/Hpa II Digested
E-ACT/ Hm-1
E-ACT/ Hm-2
E-ACT/ Hm-3
E-ACT/ Hm-4
E-ACT/ Hm-5
1
6
61
700
650
600
565
530
500- 504
460
400
364
350
300
255
200-204
145
100


Methylation-2-800
EcoR I/Msp I Digested
EcoR I/Hpa II Digested
E-ACT/ Hm-6
E-ACG/ Hm-1
E-ACG/ Hm-2
E-ACG/ Hm-3
E-ACG/ Hm-4
E-ACT/ Hm-6
E-ACG/ Hm-1
E-ACG/ Hm-2
E-ACG/ Hm-3
E-ACG/ Hm-4
1----------6 1----------6 1----------6 1----------6 1----------6 1----------6 1----------6 1----------6 1----------6 1----------6
700
650
600
565
530
500-504
460
400
364
350
300
255
200-204
145
100


Methylation-3-800
EcoR I/Msp I Digested
EcoR I/Hpa II Digested
E-ACG/ Hm-5
E-ACG/ Hm-6
E-AGC/ Hm-1
E-AGC/ Hm-2
E-AGC/ Hm-3
E-ACG/ Hm-5
E-ACG/ Hm-6
E-AGC/ Hm-1
E-AGC/ Hm-2
E-AGC/ Hm-3
1----------6 1----------6 1----------6 1----------6 1----------6 1----------6 1----------6 1----------6 1----------6 1----------6
700
650
600
565
530
500- 504
460
400
364
350
300
255
200-204
145
100


Methylation-4-800
EcoR I/Msp I Digested
EcoR I/Hpa II Digested
E-AGC/ Hm-4
E-AGC/ Hm-5
E-AGC/ Hm-6
E-AGG/ Hm-1
E-AGG/ Hm-2
1----------6
700
650
600
565
530
500- 504
460
400
364
350
300
255
200-204
145
100


Methylation-5-800
EcoR I/Msp I Digested
EcoR I/Hpa II Digested
E-AGG/ Hm-3
E-AGG/ Hm-4
E-AGG/ Hm-5
E-AGG/ Hm-6
1
61
700
650
600
565
530
500- 504
460
400
364
350
300
255
200-204
145
100