UNIVERSITY OF CALIFORNIA
RIVERSIDE

Genomic Differences Between W. Murcott Mandarin and Its Mutational Derivative Tango

A Dissertation submitted in partial satisfaction
of the requirements for the degree of

Doctor of Philosophy

in

Plant Biology

by

Yi Zhu

June 2018

Dissertation Committee:
Dr. Mikeal Roose, Chairperson
Dr. Thomas Girke
Dr. Patricia Springer

The Dissertation of Yi Zhu is approved:

_____

_____

_____

Committee Chairperson

University of California, Riverside

ACKNOWLEDGEMENTS

I would like to acknowledge all people who supported and devoted their effort to my research and dissertation. First of all, I truly appreciate Dr. Mikeal Roose for his long-term support and patient guidance throughout this dissertation project. I give many thanks to Dr. Claire Federici for providing technical supports and experimental materials that are related to my work, and managing the laboratory efficiently. I also would like to express my gratitude to my other dissertation committee members, Dr. Thomas Girke and Dr. Patricia Springer, for contributing considerable effort and time on my work during the whole process. I would like to thank Dr. Chandrika Ramadugu and Dr. Manjunath Keremane for their support on the use and data interpretation of Applied Biosystem HRM qPCR and CFX96 Real-Time PCR system, and extend many thanks to Dr. Chandrika Ramadugu for her generous assistance with DNA sequence and clones needed in my project. I am very thankful to Dr. Manuel Talon from IVIA for providing mutation information on the sequences, from which I could have a good start of this project. I also would like to thank Dr. Tim Williams and Dr. Jennifer Crowley for their previous valuable studies on Tango mandarin, based on which my project could proceed. Last, but not the least, I am grateful for my family's support, as well as encouragement from my friends, Yuanhang Huang and Lu Yu, in the most difficult period. I would have given up halfway without you.

DEDICATION

This work is dedicated to my parents for their unending love and support.

ABSTRACT OF THE DISSERTATION

Genomic Differences Between W. Murcott Mandarin and
Its Mutational Derivative Tango

by

Yi Zhu

Doctor of Philosophy, Graduate Program in Plant Biology
University of California, Riverside, June 2018
Dr. Mikeal Roose, Chairperson

The history of recorded citrus can be traced back to prior to 800 B.C. and citrus species and related genera are considered to be from the tropical and subtropical region from south-east Asia to north-eastern India, and southern China, Indo-Chinese peninsula and Malay Archipelago, and later spread to the Middle East by Arab traders and then the rest of the world. As one of the valuable consumer products, citrus has attracted great attention from researchers. Citrus breeding can create and/or improve citrus varieties using a variety of approaches, including cross-breeding, genetic transformation, mutation breeding, etc. Mutation breeding accelerates the process of mutation and provides a great opportunity to develop new traits of citrus.

The purpose of this study is to identify the genomic differences between a very low-seeded citrus cultivar, Tango mandarin, and its seedy predecessor cultivar, W. Murcott mandarin.

Tango mandarin was developed by gamma-irradiation mutation breeding of W. Murcott budwood. Both varieties share the same characteristics of size, color, sweetness and easiness of peeling, and they differ mainly in Tango having much lower pollen viability and in fruit seed content. W. Murcott fruit will be seedy unless grown in an isolated condition without cross-pollination while Tango is very low-seeded in all conditions. The intents of this study are to determine whether Tango is genetically uniform or has different genetic compositions within different cell layers, and to better characterize the putative cause of seedlessness of Tango fruit.

To determine the underlying genomic differences between W. Murcott and Tango, Illumina next-generation sequence (NGS) data of both W. Murcott and Tango leaf DNA was primarily used in analysis. Sequence data was aligned to the haploid Clementine reference genome and SNPs and Indels were predicted with the Pindel and Dindel programs. Allele-specific PCR, high-resolution melting real-time PCR and TA-cloning followed by sequencing were used for verification. For chromosome rearrangement prediction, Illumina NGS data was analyzed by novoBreak program. Copy number variation was determined by analyzing the depth of coverage of Illumina NGS reads and by loss of heterozygosity analysis with *Axiom® Citrus 15AX* SNP array data. PCR with primers flanking the putative breakpoints followed by Sanger sequencing was used to verify the chromosome rearrangements and determine the position of chromosome rearrangement breakpoints. For chimerism analysis of Tango, Tango pollen and fruit albedo were used to represent cell layer II and juice vesicles were used to represent cell layer I. Previously

verified short deletions and chromosome rearrangements were tested with DNA from these tissues to determine their genetic composition.

These experiments identified and/or confirmed three SNPs and seven Indels which were heterozygous in Tango while they were homozygous in W. Murcott. These SNPs and Indels were developed into molecular markers that differentiate Tango and W. Murcott. The copy number variation results showed no large insertions or deletions present while one heterozygous translocation of a 6 Mb segment on chromosome 2 and one heterozygous inversion of 4 Mb on chromosome 4 in Tango were verified, and their breakpoints were determined. A potential translocation between chromosome 3 and chromosome 1 was also found. These two chromosome rearrangements could be responsible for misalignment and abnormal segregation of chromosomes during meiosis as previously observed, and the resulting inviable gametes could reduce seed content of fruits. Tango pollen and fruit tissue analysis proved that two short deletions and all chromosome rearrangements were present in cell layer II but absent in cell layer I of Tango meristem, while the other five short deletions may be present in both cell layer I and cell layer II. This suggested that cell layer II, which gives rise to gametes, was affected by the initial gamma irradiation, and inviable female gametes give rise to seedless Tango fruit and inviable male gametes greatly reduce male fertility.

Table of Contents

List of Figures

List of Tables

Chapter 1: Introduction to Dissertation

Citrus History

The oldest known citrus reference is in the Sanskrit literature. A collection of texts in the Vajasaneyi Samhita dated prior to 800 B.C. referred to White Yahir-veda, which is the name applied to citron and lemon as jambhila (SCORA, 1975). Also in China, the oldest reference to citrus is in the book "Tribute of Yu". In this book, citrus was on the list of tribute items that were sent to the Emperor Yu and God of irrigation. Moreover, small mandarins such as Yuzu were sent as tributes along with other forest items by wild tribes from central and southern China (SCORA, 1975). Both in Lü's Spring and Autumn Annals in the third century B.C. and Confucius in the fifth century mentioned two citruses (SCORA, 1975). In the poem cycle Chiu Chang, the part of the Ch'u Tzu "Songs of the state of Ch'u", citrus was described in detail in the eighth poem, in which a youth was praised through comparison with a mandarin tree which is described fully in very flattering terms (SCORA, 1975). The oldest known monograph on citriculture is the book *Citrus Fruit Records* written by Han Yen-Chih in 1178 A. D., in which he named and described 27 varieties of sweet-sour orange-mandarin group in detail, as well as the distribution, cultivation and citrus orchard management (SPIEGEL-ROY AND GOLDSCHMIDT, 1996). No citrus descriptions are found in records from the ancient cultures of Egypt, Sumer and Assyria (SCORA, 1975).

The history and origin of citrus is still a controversy. Citrus species and related genera are considered to come from the tropical and subtropical region from south-east Asia to north-eastern India, and southern China, Indo-Chinese peninsula and Malay Archipelago (NICOLOSI, 2007). Gmitter and Hu considered Yunnan province to be the major center origin of citrus (GMITTER AND HU, 1990), while some recent studies indicate that the true origins of citrus are from Australia, New Caledonia and New Guinea (LIU et al., 2012). Beattie et al. also hypothesized that *Citrus* genus was dispersed to the westward from Australia to South Asia (BEATTIE et al., 2008). Later citrus was brought to the Middle East by Arab traders and then spread to the rest of the world slowly as a consequence of travel and exploration, including northern Africa and southern Europe (ENSMINGER AND ENSMINGER, 1993; LIU et al., 2012). The first citrus seeds were brought to Haiti by Columbus on his second voyage in 1493, followed by the orange reached Mexico in 1518 (SCORA, 1975). In the United States, citrus was brought to St. Augustine, Florida by Spanish settlers in 1565, and spread to Arizona in the eighteenth century, and Citrus orchards were first developed in Florida and California around seventeenth century and eighteenth century (LIU et al., 2012; SCORA, 1975).

Citron (*Citrus medica* L.) is native to India and the first citrus fruit known to Europeans (NICOLOSI, 2007). When it was dispersed over the Hellenistic Near East, it became an essential part of the Jewish feast of Tabernacles. Citron was also traded over the entire Mediterranean region through the Jewish communities (SCORA, 1975). Due to its sanctity, pleasant odor, good appearance and relative imperishability during long-time travel, citron became the first citrus reaching the West and attracted the attention of Europeans, while

sour oranges, lemons and sweet oranges reached Europe several centuries later (SCORA, 1975). Citron is also the first described species of citrus and it is considered to be from India (BAYER et al., 2009). In India, citron was sanctified and consecrated to the God of Wisdom (SCORA, 1975). Lemons and oranges were taken to northern California from Baja California by the Franciscan missionaries (SCORA, 1975).

Sour orange (*Citrus aurantium* L.) originated from south-east Asia (NICOLOSI, 2007; WEBBER AND BATCHELOR, 1943). The first description of sour orange was made by Albertus Magnus, in which sour orange was called *Arangus* (NICOLOSI, 2007).

Sweet orange [*C. sinensis* (L.) Osbeck] is native to southern China and maybe as far south as Indonesia (WEBBER AND BATCHELOR, 1943). It became well-established and commercially valuable in Southern Europe by the beginning of the sixteenth century (NICOLOSI, 2007). Orange, lime and lemon trees were taken to Port Jackson in Australia from Rio de Janeiro in 1788 by the settlers of the first English fleet sailing (SCORA, 1975). The first sweet orange trees in Holland were supplied by the island of Saint Helena, and then sweet oranges were brought and planted in Capetown when the Dutch colonized South Africa (SCORA, 1975).

Mandarin (*Citrus reticulata* Blanco), which originated from northern India or southern China, was first brought to England, and spread to Malta, Sicily and Italy in the nineteenth century (SCORA, 1975; SWINGLE, 1967). Mandarin has been cultivated since about 4000 years ago. Based on color, shape, surface and size of the fruit, thickness and smell of the peel, taste of the pulp and morphological characteristics of the tree, the mandarins in China

were originally divided into two groups: *Gan* (*Macroacrumen*) and *Ju* (*Microacrumen*) (LI et al., 1992). When mandarins were cultivated in China as early as 21st-16th century BC, there was no concept of citrus varieties. The word *Ju* was used for all citrus plants, including sour orange and the Yuzu (*C. junos* Tanaka) (NICOLOSI, 2007). In Japan, mandarin was brought from China and became the foremost citrus.

Pummelo (*Citrus grandis* Osbeck) is from the Malayan and East Indian archipelagos, and spread to China, then Persia, Palestine and Europe (SPIEGEL-ROY AND GOLDSCHMIDT, 1996). It has been commercially grown in the East since it was introduced to Japan in the 1770s (NICOLOSI, 2007). The pummelo was first described and named 'shaddock' by Sloane in Jamaica (WEBBER, 1943). Grapefruit, a natural hybrid of pummelo, is grown mainly in America (SAWAMURA AND KURIYAMA, 1988). Pummelo fruits have a pleasant taste with sugar and acid and a refreshing flavor like that in the grapefruit (SAWAMURA AND KURIYAMA, 1988). Both sour orange and sweet orange are produced by introgression of predominantly mandarin genotype with genes from pummelo (MOORE, 2001).

Most recent research supported the idea that citrus originated from southeast Asia, specifically in western Yunnan province of China, northern Myanmar and northeastern India in the Himalayan foothills, and it diversified and radiated during the late Miocene epoch (WU et al., 2018). Later the climate transition caused the migration of citrus across the Wallace line and formation of the second radiation in Australasia during the early Pliocene epoch (WU et al., 2018).

Citrus Taxonomy

The broad morphological diversity and polyembryony of citrus, in which many maternally derived embryos can develop from the nucellus together with the zygotic embryo, can account for the complexity of citrus taxonomy (GARCIA-LOR et al., 2012; KOLTUNOW et al., 1996). Additional contributing factors to the taxonomic controversy are the interspecific compatibility within the genus and inclination to bud mutation (GARCIA-LOR et al., 2012; LIANG et al., 2007). In previous studies, morphological characteristics and secondary metabolites demonstrated that the majority of edible citrus are generated from the differentiation between three basic taxa: mandarin, pummelo and citron (GARCIA-LOR et al., 2012). Although the first description of morphology and uses of the trees, flowers and fruits date back to ancient times, the first description and classification of citrus species and varieties was completed in the 17th century by John Baptista Ferrarius in this book *Hesperides, sive de Malorum Aureorum culturea et usu* (NICOLOSI, 2007).

It is accepted by most taxonomists that citrus species belong to the Geraniales order, the Rutaceae family and the Aurantioideae subfamily (NICOLOSI, 2007). Swingle described the Citrus morphology and that of its relatives (SWINGLE, 1967). Citrus taxa were also reclassified by Mabberley (MABBERLEY, 2004). Recent molecular studies estimated that the age of Rutaceae is 91 million years, the Aurantioideae is 71 million years and *C. trifoliata* auth. (*Poncirus trifoliata*) is 18 million years (BAYER et al., 2009). The Citrus genus was considered to consist of two major clades, Australasian clade and Asiatic clade (BAYER et al., 2009). Australasian clade consists of *Clymenia* and *Microcitrus* in Papua New Guinea, *Eremocitrus* and *Microcitrus* in Australia, and *Oxanthera* in New Caledonia

(CURK et al., 2016). The Asiatic clade consists of the well-known commercial citrus species and hybrids as well as *Fortunella* and *Poncirus* (BAYER et al., 2009).

Swingle subdivided the Aurantioideae into two tribes: Clauseneae with 5 genera and Citreae with 28 genera including *Citrus* and related genera (*Fortunella*, *Poncirus*, *Eremocitrus*, *Microcitrus* and *Clymenia*) (SWINGLE, 1967). The Clauseneae tribe consists of more primitive genera: Micromelinae, Clauseneae and Merrillinae (SWINGLE, 1967). The fruits are usually small with semi-dry to juicy berries. The Citreae tribe consists of three subtribes: Triphasiinae, Citrinae and Balsamocitrinae (SWINGLE, 1967). The classification of tribes, subtribes, genera and species within the Aurantioideae subfamily of Rutaceae family is the most controversial, because Citrus and many related genera are easy to hybridize.

Two taxonomic schemes are broadly accepted: the "wide species" classification by Swingle (SWINGLE, 1967) and the "narrow species" classification by Tanaka (LIANG et al., 2007; TANAKA AND TANAKA, 1954). Based on their morphological characteristics and chemical components of flowers, leaves and fruits, Swingle's divided the *Citrus* genus into two subgenera: *Citrus*, including ten species, and *Papeda*, including six species (NICOLOSI, 2007). While Tanaka divided the *Citrus* genus into two subgenera: *Archicitrus* and *Metacitrus*, including 157 species in total (TANAKA, 1961). Although Tanaka system is more complicated and provides an exhaustive description of citrus taxa, Swingle system is the most used due to its being easier to understand (NICOLOSI, 2007). Other classification systems also exist, such as Marcovitch's classification which divided the *Citrus* genus into three groups: *Aurantium*, *Intermedium* and *Medica*, and 20 species (MARCOVITCH, 1926).

The taxonomic classification of mandarins is the big difference between the Swingle and Tanaka systems. In the Swingle system, all the mandarins except *C. tachibana* and *C. indica* are included in the species *Citrus reticulata*, while in Tanaka system, mandarins are separated into 36 species (NICOLOSI, 2007). The Citrus classification used to be performed by morphological characters, isozyme analysis, cytology, paleontology and stoichiometric chemistry, and currently done by molecular markers including DNA sequences (BAYER et al., 2009; LIANG et al., 2007).

In addition to morphological classification, chemical components and biochemical properties of citrus fruits, such as fatty acids, hydrocarbon profiles, flavonoid patterns, carotenoid composition and enzymes, were also utilized to identify citrus varieties and develop citrus systems (LIU et al., 2012). Biochemical studies, numerical taxonomy and molecular studies all showed that cultivated Citrus species arise from interspecific hybridization between four basic taxa: mandarin (*C. reticulata* Blanco), pummelo (*C. maxima* (Burm.) Merrill), citron (*C. medica* L.) and papeda (*C. micrantha* Wester) (CURK et al., 2016). Other varieties are considered to be either derived through cross-hybridization amongst the basic species or cross-hybridization between them and closely related species (PANG et al., 2007). Besides *Citrus*, the subtribe *Citrinae* contains the other five closest relatives of *Citrus*: *Poncirus*, *Eremocitrus*, *Microcitrus*, *Fortunella* and *Clymenia* (SWINGLE, 1967). However, there is little evidence that any of these taxa participated in the origin of most cultivated citrus.

Mabberley had new views of citrus and the relationships between “the true citrus types” which should include *Oxanthera* and *Feroniella* in addition to Swingle’s “microgenera”,

Clymenia, *Eremocitrus*, *Fortunella*, *Microcitrus*, and *Poncirus* (BAYER et al., 2009; MABBERLEY, 1998). In his studies, the citron was involved in lemon (*Citrus* × *limon*), rough lemon (*Citrus* × *jambhiri*), lime (*Citrus* × *aurantiifolia*) and bergamot (*Citrus* × *bergamia*) and their similar hybrids. The pummelo was involved in *Citrus* × *aurantiifolia* and *Citrus* × *aurantium*, which includes three pummelo hybrids: sour orange (*Citrus* × *aurantium*), sweet orange (*Citrus* × *sinensis*) and grapefruit (*Citrus* × *paradisi*). The mandarin included mandarin, satsuma, clementine and tangerine (MABBERLEY, 1997). He suggested that there may be only 25 true-breeding species and that the first species might have dispersed from North-Eastern Australasia and spread as 'floating fruit' to westward-flowing equatorial currents around 30 million years ago (LIU et al., 2012).

C. reticulata diversified over Southern China, Vietnam and Japan, and *C. maxima* originated in the Malay Archipelago and Indonesia, and *C. medica* evolved in Northeastern India and near Burma and China (GARCIA-LOR et al., 2012). Before the migration and hybridization facilitated by human activity, the foundation effect in the three geographic zones and an initial allopatric evolution caused the differentiation between the three basic taxa (GARCIA-LOR et al., 2012). The secondary species, sour orange, sweet orange, lemon and lime, are derived from recombination among the three basic taxa (OLLITRAULT et al., 2003).

Mandarin has a broad genetic base, and it can produce zygotic and nucellar embryos. The ancestral *C. reticulata* group contributes a large proportion of the genomes of secondary species and most intervarietal polymorphisms result from point mutation or transposable elements (GARCIA-LOR et al., 2012). The diversification of several apomictic species

mainly occurred through mutation or epigenetic variations (HÖRANDL AND PAUN, 2007). According to SSR and Indel molecular marker analysis result, modern tangor and tangelo cultivars show higher heterozygosity than *C. reticulata*, *C. maxima* and *C. medica*, and positioned between the three basic taxa on marker PCA analysis, which also confirmed their highly hybrid status (GARCIA-LOR et al., 2012).

Clementine is one of the most popular mandarin varieties in the Mediterranean Basin, which was selected from a seedling of ‘‘Common mandarin’’ in Algeria one century ago, and maybe a hybridization of ‘‘Common mandarin’’ × *C. sinensis* (GARCIA-LOR et al., 2012). This marker study showed that it genome displayed most contribution from *C. reticulata* and around 6% contribution from *C. maxima* (GARCIA-LOR et al., 2012). However, later genome sequence analysis indicates the proportion from *C. maxima* is 21% (Gmitter et al., 2014). The mandarin hybrids, such as tangor and its varieties “Temple”, “Ellendale”, “Murcott” and “King”, are hybrids between *C. reticulata* and *C. sinensis*, consisting of 90% contribution from *C. reticulata* genome and 10% contribution from *C. maxima* genome (GARCIA-LOR et al., 2012). Tangelo is a hybrid between *C. reticulata* and *C. paradisi*, and has approximately 20% contribution from *C. maxima* (GARCIA-LOR et al., 2012). Possibly analysis of genome sequences will alter this proportion as occurred for Clementine.

C. medica was found entirely cytogenetically homozygous according to some molecular studies, with heteromorphisms restricted to small variations in chromosome size or band size (CARVALHO et al., 2005). The cytogenetical homozygostiy of *C. medica* may be because self-pollination of cleistogamous citron flowers produces vigorous selfed

seedlings which are highly homozygous (BARRETT AND RHODES, 1976; CURK et al., 2016). Therefore *C. medica* can be considered as a true species that has contributed to the origin of other citrus hybrids such as lemon and lime (CARVALHO et al., 2005).

For the secondary cultivated species, the precise contribution of the three basic species is not precisely known for all of them. Due to the limited number of interspecific recombinations, the high and generalized linkage disequilibrium was observed in the secondary cultivated species (GARCIA-LOR et al., 2012). Ollitrault et al. developed Indel markers and used them in genetic diversity and phylogenetic studies (GARCIA-LOR et al., 2012). Comparing to SSR markers, which is more suitable for intraspecific diversity analysis, the Indel markers are more useful as phylogenetic markers for tracing the contributions of three basic citrus species (GARCIA-LOR et al., 2012).

Sweet orange (*C. sinensis* (L.) Osb.) and sour orange (*C. aurantium* L.) are closely related to mandarin orange and contain traits and markers from pummelo. In previous studies, the sweet orange genome was inferred to consist of 75% from *C. reticulata* and 25% from *C. maxima* by using molecular markers, indicating that they are possibly hybrids generated from backcrossing of first or second female generation with the male mandarin orange ((*C. maxima* × *C. reticulata*) × *C. reticulata*), rather than direct hybrids (GARCIA-LOR et al., 2012). However, in a recent study, sweet orange was found to share alleles with Ponkan mandarin across about 75% of the genome, and many of those alleles are also shared with Willowleaf mandarin and Huanglingmiao mandarin, which indicated the kinship of the ancestor of sweet orange with the three mandarins (WU et al., 2014). Both parents of sweet orange were found to have some pummelo ancestry, and the chloroplast genome of sweet

orange also found to be pummelo-type, which indicated the female parent could be a pummelo with introgression of some wild mandarin (WU et al., 2014). Recently the release of Valencia sweet orange draft genome, whose assembled sequence covers 87.3% of the estimated orange genome with 20% repetitive elements, also showed evidence of a backcross hybrid character of sweet orange between mandarin and pummelo (XU et al., 2013).

Sour orange displays a contribution of both *C. maxima* and *C. reticulata* gene pools, consistent with it being a cross between a pummelo as the seed parent and a wild-mandarin as pollen parent (WU et al., 2014). However, in contrast to sweet orange, no pummelo admixture was found in the mandarin parent of sour orange (WU et al., 2014). Therefore, sour orange could be a natural hybrid of two wild Citrus species followed by apomictic reproduction before human cultivation (WU et al., 2014). Some earlier studies showed the approximate proportion of contributions as 68% from *C. reticulata*, 30% from *C. maxima* and 2% from *C. medica*, and therefore it can be a hybridization of backcross of C. maxima ×(C. maxima ×C. reticulata) (GARCIA-LOR et al., 2012). The studies also showed that the sour orange might derive from the same parentage as Suntara mandarin since they share the same alleles at most loci (GARCIA-LOR et al., 2012). The pummelo contribution was also found in the chloroplastic and mitochondrial genomes in sour orange (GARCIA-LOR et al., 2012).

Grapefruits (*C. paradisi* Macf.) are highly polyembryonic and genetic variation among grapefruit cultivars was found to be very low; therefore the different cultivars are likely of nucellar or bud sport origin (AYDIN UZUN et al., 2010). The molecular marker analyses

confirmed that the grapefruit is closely related to pummelo and display introgression from sweet orange, whose genome is contributed by 61% from *C. maxima* and 39% from *C. reticulata*, indicating the theoretical hybridization of *C. maxima* × [(*C. maxima* × *C. reticulata*) × *C. reticulata*] (GARCIA-LOR et al., 2012). The PCA analysis of 506 markers indicated the intermediary position of grapefruit between the sweet orange and pummelo gene pool, supported by 96.3% to 98.0% of analyzed markers (OLLITRAULT et al., 2012). The genetic composition of introgression of mandarin to pummelo was also supported by analysis of both amylase polymorphism and flavonoids, and sweet orange served as the pollen parent while pummelo served as the maternal parent for grapefruit (SCORA et al., 1982).

Citron (*C. medica*) is a progenitor of limes (*C. aurantifolia* (Christm.) Swing) and lemons (*C. limon* Osb.) (GARCIA-LOR et al., 2012). The chloroplast and nuclear data analysis showed that the genetic pools of mandarin orange and pummelo also contributed to the genesis of lemon (GARCIA-LOR et al., 2012).

Lemons and limes are the third largest citrus horticultural group being cultivated in the Mediterranean, sub-tropical and inter-tropical climates worldwide for production of fresh fruit and essential oils or use as rootstock (CURK et al., 2016). Most lime and lemon prototypes originated in Asia and the diversification areas of the presumed parents overlap (CURK et al., 2016). Limes and lemons are highly heterozygous and highly polymorphic with diploid, triploid and tetraploid varieties (CURK et al., 2016). All limes and lemons contain a contribution from *C. medica* genome as a direct male parent, and female parents include *C. micrantha* or a closely related papeda species, *C. reticulata*, *C. aurantium* or *C.*

maxima × *C. reticulata* hybrids (CURK et al., 2016). All citron chromosome types are clearly represented in lemons and limes and some form heteromorphic pairs (CURK et al., 2016). Both 'Tahiti' and 'Persian' lime belong to *C. latifolia* triploid accessions resulting from a haploid ovule of *C. limon* fertilized by a diploid gamete of *C. aurantifolia* (CURK et al., 2016). 'Tanepao' lime and 'Madagascar' lemon belongs to *C. aurantifolia* triploid accessions and their being triploid was caused by an interspecific backcross involving fertilization of a diploid ovule of *C. aurantifolia* by *C. medica* (CURK et al., 2016). Lemon cv. 'Eureka' and 'Lisbon' are tri-hybrids consisting of 41% from *C. reticulata*, 45% from *C. medica* and 13% from *C. maxima,* while Mexican lime is hybridized from citron and a Papeda species and *C. micrantha* might be the parental Papeda (BARRETT AND RHODES, 1976; CURK et al., 2015; CURK et al., 2016; FEDERICI et al., 1998; GARCIA-LOR et al., 2013; GARCIA-LOR et al., 2012; NICOLOSI et al., 2000; OLLITRAULT et al., 2012).

W. Murcott Mandarin and Tango Mandarin

W. Murcott mandarin, also known as Nadorcott or Afourer, is originally from Morocco and was imported to California in 1985 (OKIE, 2000; SAUNT, 1990). It is believed that W. Murcott is a complex hybrid of a chance zygotic seedling from the Murcott tangor and an uncertain pollen parent, where Murcott itself is presumed an F1 hybrid from a cross between an unknown mandarin and a sweet orange (WU et al., 2014). The sequence analysis indicated 34% of sweet orange genome is shared with W. Murcott, supporting the suspected grandparent-grandchild relationship between sweet orange and W. Murcott (WU

et al., 2014). W. Murcott fruits have deep orange color, are sweet and juicy, and they are easily peelable mid-late season mandarins (ROOSE AND WILLIAMS, 2007a). Because of their high quality, they are widely planted throughout California. Moreover, a significant feature is they can be seedless, because W. Murcott is self-incompatible, which means the male gametes are unable to fertilize female gametes, preventing seed formation and resulting in facultative parthenocarpy (GAMBETTA et al., 2013; OLLITRAULT et al., 2007). When W. Murcott mandarin trees are being planted, growers either plant them in isolated areas or cover whole rows of W. Murcott mandarin trees with the screen during flowering season to prevent cross-pollination (CROWLEY, 2011). Otherwise, they will produce seedy fruits as a result of cross-pollination by bees-transferring pollen from other Citrus with high pollen viability, which could significantly reduce the fruit value (CROWLEY, 2011).

Because it is time-consuming and labor-consuming to isolate W. Murcott mandarin trees from other citrus trees to prevent cross-pollination, the UCR breeding program developed a new cultivar from W. Murcott that is seedless even in cross-pollination conditions (ROOSE AND WILLIAMS, 2007a). Ionizing radiation is a conventional technique to induce mutations in plant breeding. High levels of radiation on citrus can induce genetic alterations and result in citrus cultivars bearing fewer or no seeds (HEARN, 1984; ROOSE AND WILLIAMS, 2007a; VARDI et al., 1996). Budwood is collected from a citrus tree and exposed to the desired level of ionizing radiation. The shoots which develop from the cells of meristem in the irradiated budwood are propagated into trees among which trees bearing seedless fruits can sometimes be found (YE et al., 2009).

Tango is derived from gamma ray irradiation mutation breeding of budwood of the diploid W. Murcott mandarin cultivar (ROOSE AND WILLIAMS, 2007a). After budwood of W. Murcott was treated with gamma rays to induce mutations, the irradiated buds were propagated and the resulting trees evaluated. The most promising selection was eventually named Tango and released. Tango is similar to W. Murcott for it also has deep orange rind color and sweet juicy fruit. In contrast with W. Murcott mandarin, Tango has relatively low seed count (0.2 seeds/fruit), as well as low pollen viability and low pollen germination rate (<5%) (ROOSE AND WILLIAMS, 2007a). Hand pollination trials of Clementine mandarins with Tango pollen did not cause Clementines to bear seedy fruit (ROOSE AND WILLIAMS, 2007a). Also, the seed content of Tango mandarin is little affected by pollen from seed bearing citrus. When grown with cross-pollination, Tango mandarin typically have less than one seed/fruit, while W. Murcott mandarin usually has 11-22 seeds/fruit (ROOSE AND WILLIAMS, 2007a).

Seedless Fruit

Fruit development is closely associated with seeds, and the related processes are controlled by internal signals and environment (PANDOLFINI, 2009). The seedless fruits are appreciated by consumers and producers in both fresh consumption and conserved or processed fruits for the seedlessness increases the fruit quality and shelf-life (PANDOLFINI, 2009). Usually, the substances the seeds produce accelerate the deterioration and senescence of the fruit, so the absence of seeds can provide better preservation

(PANDOLFINI, 2009). For example, the seedlessness of eggplant can prevent browning and texture reduction of the pulp (PANDOLFINI, 2009).

Classical breeding methods for seedless fruit are commonly based on hybridization between seeded and seedless varieties, but the drawback is that hybridization relies on the choice of parents and the proportion of seedless plants in the progenies is often relatively low (BOUQUET AND DANGLOT, 1996). Also, the long juvenile period after the cross has been made before fruits are produced consistently and inbreeding depression make the conventional approaches difficult (PERL et al., 1998).

Seedless fruit can be produced at different fruit development phases. During the ovary development, fertilization and fruit setting phase, the ovary mutant (such as Corinth grapes, cucumber and tomato), incompatible pollen (such as citrus fruits) and sterile pollen (such as cucumber) can lead to the failure of fertilization (VAROQUAUX et al., 2000). In some plants, the ovary is capable of developing even without pollination or ovule fertilization and become seedless fruit, the process of which is called parthenocarpy (VAROQUAUX et al., 2000). Parthenocarpy can occur naturally or artificially by application of plant hormones (GORGUET et al., 2005). During the fruit cell division, embryo development and seed formation phase, abnormal meiosis in the triploid plant (such as watermelon) and defects in endosperm development (such as grapes) can also lead to seedless fruits, which is called stenospermocarpy (VAROQUAUX et al., 2000).

In grape, there are generally two types of seedless fruit: parthenocarpic cultivars such as Black Corinth in which seeds are never formed, and stenospermocarpic cultivars such as

Sultanina in which embryo and/or endosperm abort and result in rudimentary seeds (BURGER et al., 2009). No seedless progeny could be obtained when parthenocarpic cultivars were used to pollinate various seeded cultivars, while when stenospermocarpic cultivars were used as pollinators of seeded cultivars, seedless progeny could be obtained (BURGER et al., 2009).

In ovulo and *in vitro* culture techniques to rescue viable embryos are widely applied to the breeding of seedless fruits (BOUQUET AND DANGLOT, 1996). These methods make possible the recovery of a higher percentage of seedless plants from progenies of crosses seedless by seedless (BOUQUET AND DANGLOT, 1996).

Seedless fruit can be obtained by mutation occurring within specific genes responsible for specific steps in reproduction without altering the entire genetic makeup of a genotype (KHAWALE et al., 2004). A partially sterile form, which has the potential for commercial adaptation, of a seedless Perlette was obtained by mutation breeding (KHAWALE et al., 2004).

Another method to produce seedless fruit is using phytohormones, such as auxin, gibberellin, cytokinins or hormone mixtures, which control the initiation and regulation of fruit growth (PANDOLFINI, 2009). The widespread practice is to treat flowers with phytohormones before pollination to produce seedless parthenocarpic fruit (PANDOLFINI, 2009). The increased level of phytohormones in the ovary can substitute for pollination and parthenocarpic fruit development is thus triggered (GORGUET et al., 2005). With this method, the fruit can be produced independently from pollination and fertilization of the

ovules (PANDOLFINI, 2009). Tissue-specific hormone manipulation is also used to avoid abnormal growth phenotypes caused by exogenous phytohormone applications, which may influence too many processes in plant growth and development (GORGUET et al., 2005).

The genetic method is to transfer individual traits in the form of single genes into the background that is already available and desirable with minimum genome disturbance (PERL et al., 1998). Its application for seedless fruit is valuable if the quality and productivity remain the same except for the lack of seeds (PANDOLFINI, 2009). Seedless mutants with that have hormone biosynthesis or hormone perception genes altered are broadly studied (GORGUET et al., 2005). The choice of promoter should be genes that express specifically in ovules, ovaries or other fruit tissues during a development stage unique to fruit (PANDOLFINI, 2009). For example, in tomato, eggplant and tobacco, the gene expression of *DefH9-iaaM* chimeric gene in ovules, placenta and tissues derived from them allows auxin synthesis in later developmental stages of fruit growth, leading to seedless parthenocarpic fruits from emasculated flowers, fruits equal or larger in size compared to those of pollinated flowers without changing skin color, flesh consistency or percentage of fruit set (FICCADENTI et al., 1999; PANDOLFINI, 2009). Another example is the spindly (*spy*) mutant of *Arabidopsis*, in which parthenocarpic fruit is produced due to hormonal deregulation (JACOBSEN AND OLSZEWSKI, 1993). Also, facultative parthenocarpy can be controlled genetically to make plants set seedless fruit under pollination-restrictive conditions (CARMI et al., 2003). Seedless fruit can also be obtained by *Agrobacterium*-mediated genetic transformation of embryogenic cell lines followed by expression of a gene controlling seed development (PERL et al., 1998).

In grape, QTL studies discovered linked genes that control the stenospermocarpic seedlessness (DOLIGEZ et al., 2002). For example, a genetic model was confirmed that the expression of seedlessness was controlled by a dominant *sdI* (seed development Inhibitor) gene and some other complementary recessive genes that are regulated by it (LAHOGUE et al., 1998). Other molecular markers were also found linked to the loci that are involved in seedless trait and developed for the selection of seedless progenies (ADAM-BLONDON et al., 2001). Also by using the seedless genotypes as female parents combined with in vitro embryo rescue, a significant proportion of entirely seedless genotypes can be generated through seedless × seedless crosses (DOLIGEZ et al., 2002).

In citrus, the seedlessness of fruits is an attractive character to consumers and advantageous for the juice industry. To develop seedless citrus cultivars has been a critical breeding objective for citrus breeders. Most triploid seedless citrus varieties emerge spontaneously or by diploid-tetraploid crosses, and a small proportion of are unexpected triploids from diploid parents by the production of diploid egg cells or both diploid egg cells and diploid pollen grains uniting with monoploid gametes (ESEN AND SOOST, 1971). It is also common in citrus that gametophytic self-incompatibility coupled with parthenocarpy causes seedlessness by inhibiting the pollen tube growth in the style or ovary and so blocking fertilization (YE et al., 2009; ZHANG et al., 2012). The self-incompatibility is an essential character of many plants controlled by a single multi-allelic locus (the S locus), preventing self-fertilization and encouraging outcrossing (YE et al., 2009). In contrast, some self-compatible citrus varieties, such as Zigui Shatian pummelo, are seedless even though they are successful in self-pollination and fertilization because of they have self-incongruity, in

which seeds are aborted due to abnormal post-zygotic development (CHAI et al., 2011). Cytoplasmic male sterility, which inhibits the generation of functional pollen, can produce seedless fruits, such as in the naturally seedless (or low-seeded) variety Satsuma mandarin (CAI et al., 2007). Moreover, transferring cytoplasmic male sterility to seedy citrus cultivars by somatic hybridization offers a novel strategy for breeding of new seedless diploid citrus (CAI et al., 2007). Reduced pollen fertility and increased ovule sterility caused by mutation are frequently observed in seedless and low-seeded citrus, and other seedless mechanisms including embryo sac abortion and chromosome ploidy variation are also observed (DENG et al., 1996; HEARN, 1984; HEARN, 1986; SPIEGEL-ROY AND VARDI, 1990). Anther abortion during the initiation of anther development that results in no pollen grain produced can arise from gene-cytoplasmic interaction (YAMAMOTO et al., 1997). In some citrus cultivars such as Huami Wuhegonggan, seedlessness is due to both male sterility and embryo abortion after self- and cross-pollination, where the latter is the primary cause (QIN et al., 2015). In male sterile cultivars, seedless fruits can be obtained by parthenocarpy coupling with hormone application such as gibberellin (TALON et al., 1992). Female sterility is another mechanism responsible for stable seedless fruits of some citrus. In the studies of Lipeng No.2 Ponkan, the embryo sac mother cells were found almost degenerated during megasporogenesis and produced seedless fruit (XIAO et al., 2007). The senescence or non-formation of the embryo sac of URS Campestre seedless orange leads to the absence of the female gametophyte, impeding fertilization and seed development (BENDER et al., 2017). Embryo abortion, abnormal embryo sacs and unfertilized ovules can also cause seedlessness (QIN et al., 2015). Polyploids, particularly

triploids and tetraploids, which are common to *Citrus* and related genera such as *Poncirus* and *Fortunella*, can be generated either spontaneously or by certain crosses and are a conventional method for seedless citrus breeding (ESEN AND SOOST, 1972; ESEN et al., 1978). Triploid citrus is hugely desirable to breeders because they can produce seedless (or low seeded) fruits (ESEN et al., 1978). Triploids can be produced through crosses between diploid and tetraploid, and 4X ×2X crosses had been proved to be more efficient to produce triploid seed than 2X ×4X crosses in several studies due to the degeneration or subnormal development of the 4X endosperm (ESEN AND SOOST, 1973; ESEN et al., 1978). Also, both seed size and weight of triploid embryos during either embryogenesis or seed maturation were significantly reduced in comparison to those of diploids (ESEN AND SOOST, 1971). 'Oroblanco' and 'Melogold' are such seedless triploids of the hybrid cross between pummelo and tetraploid grapefruit (SOOST AND CAMERON, 1985; SOOST AND CAMERON, 1980). Triploid citrus varieties can also be produced by hybridization of a diploid pollen parent and a female parent that is obtained from regenerated somatic hybrids between two favorable monoembryonic diploid cultivars (GUO et al., 2000). Diploid megagametophytes are also used for generation of seedless varieties, such as Temple, Wilking and Fortune mandarins (CUENCA et al., 2011).

Mutation Breeding in Citrus

Although conventional crossbreeding has been successfully applied in many crops, the long juvenile period, heterozygosity, polyploidy and high space-consuming have become

the limitations for fruit trees (SPIEGEL-ROY, 1990). Instead, mutation breeding has taken a significant part in new cultivar development, including improvement of fruit color, fruit quality, longer time periods to fruit maturity, and seedlessness (VARDI et al., 2008). While natural mutants have been widely accepted by the public, a shorter period for obtaining new cultivars than is typically possible when relying on natural mutants is needed to satisfy market requirements. Induced mutation has been developed to overcome the low frequency of spontaneous mutation in order to accelerate the process of developing new traits and/or seedless fruit from existing commercial cultivars, and the mutated germplasm can be used as a parent in a hybridization selection program (ROOSE AND WILLIAMS, 2007b; VARDI et al., 2008). Also, the characteristics of a high degree of apomixis and polyembryony in citrus impede efficient conventional breeding to produce intra-specific hybrids, and therefore mutation breeding provides a great opportunity when conventional breeding methods are constrained (HENSZ, 1991). Therefore, besides embryo sac abortion, odd polyploidy and self-incompatibility, seedless citrus cultivars have been achieved by seed or budwood exposure to neutron irradiation, X-ray, gamma ray, etc. that cause ovule or pollen sterility (BENDER et al., 2017; BERMEJO et al., 2011; DENG et al., 1996). In some work, seedless citrus fruits were obtained by treating apomictic seeds with gamma rays and thermal neutrons (SPIEGEL-ROY, 1990). Thermal neutron treatment of seeds and budwood has been applied to induce mutation by which seedless fruit, redder flesh and peel color cultivars have been generated (HENSZ, 1991). Ionizing radiation has been broadly utilized since it often affects more than one gene by inducing many non-lethal chromosome aberrations, while the drawback is pleiotropic effects and unexpected linkage of genes

(SPIEGEL-ROY, 1990). Somaclonal variation and transgenic approaches can also induce new mutants but have resulted in few commercial cultivars (SPIEGEL-ROY, 1990). Chimera formation and consequent potential for reversion can occur with irradiation of meristematic multicellular buds; therefore the selection of second vegetative generation progeny for desired traits and further clonal trials are usually performed to reduce the frequency of these problems (SPIEGEL-ROY, 1990).

Since the economic value and marketability of citrus are associated with the number of seeds per fruit, and natural seedless mutants of great significance have been found in various citrus species, mutation breeding has been emphasized in citrus. Spiegel-Roy and Vardi developed budwood irradiation of citrus, in which the buds were exposed to ^{60}Co gamma irradiation followed by grafting to rootstock for further observation and selection (GIDONI AND CARMI, 2007). Different doses of gamma rays have been used to irradiate buds with the aim of producing seedless or low-seeded fruits in breeding programs (VARDI et al., 2008). By utilizing budwood irradiation with gamma rays, several seedless citrus mutants from seeded parents, including 'Eureka' lemon, 'Villafranca' lemon and 'Minneola' tangelo, have been produced, in which the ovules were almost completely sterile while pollen was fertile (SPIEGEL-ROY AND VARDI, 1990). 'Eureka' and 'Villafranca' were reported completely seedless and without reversion, while average seed count of 'Minneola' was reduced to 0.6 seeds per fruit (SPIEGEL-ROY, 1990). Following treatment with X-rays and thermal neutrons, 'Star Ruby' and 'Rio Red' were commercially released as new cultivars with three times redder flesh color than most natural pink grapefruit mutants, and lower seed counts than their parent, 'Hudson Red' (HENSZ, 1991).

‘Mor’ mandarin is a low-seeded cultivar with low pollen fertility that was developed from mutation breeding from ‘Murcott’ (VARDI et al., 1993). Other studies also found that pollination with irradiated sterile pollen could also prompt the development of seedless fruit on both self-incompatible and self-compatible cultivars, and the pollen fertility may be correlated with seed number per fruit (VARDI et al., 1995). In studies of the seedless bud sport mutant ‘Ougan’ mandarin (*C. suavissima*), pollen was found aborted at the tetrad stage of microspore development in anthers, with heterogeneous microspore populations in the same microsporangium and almost no normal tetrads observed (HU et al., 2007). While being grown alongside other citrus cultivars, seedless ‘Ougan’ mandarin became seedy, which indicated the female gametes were fertile and self-pollination with aborted pollen was responsible for seedlessness (HU et al., 2007). Therefore, the selection of mutants with both male and female sterility is preferred, to avoid the pollination restrictions. The sparse-seeded mutant Kinnow (*Citrus reticulata* Blanco) from budwood irradiation was found to be caused by both ovule abortion and following seed abortion, and might involve a recessive gene or genes that affects the chemicals functioning in the termination of the seed development process (KHALIL et al., 2011). A gene was found to be related to asynapsis in the inbred progeny of cv. ‘Wilking’ (SPIEGEL-ROY AND VARDI, 1990). Recent combination of advanced sequencing and molecular biology techniques empowers faster and more efficient approaches for citrus mutation breeding than random mutagenesis, which requires screening of large populations for the desirable traits (GIDONI AND CARMI, 2007).

Mutation breeding with somatic tissue is prone to chimera formation because different mutations can be induced in different cells within the same apex along with non-mutant cells in multicellular apices (VARDI et al., 2008). The formation of chimera could be a challenge and have a negative attribute for both growers and consumers. Chimerism in citrus was reported in the studies of Shamel et al. in 1925, where two kinds of fruit rind were found present in the same fruit of some "dual" orange strains and those strains could be periclinal chimeras (SHAMEL et al., 1925). Later the somatic chromosome duplication was found on one of the progenies of a hybrid mandarin, King × Dancy, which produced buds of at least two types of cytological constitution: 2n-4n-4n and 2n-4n-2n, corresponding to polyploidy in the second and third or second only cell layers of the meristem respectively (FROST AND KRUG, 1942). The layers with different ploidies also suggested the participation of all three cell layers in vegetative shoot development and the resulting periclinal chimera, based on morphology and chromosome counts of tissues derived from them (FROST AND KRUG, 1942). Also, tetraploidy and albino of fruit sector chimeras found in sweet oranges, grapefruit, tangelo and tangors indicated the occurrence of genetic mutations (BOWMAN et al., 1991). Sectored fruit chimeras caused by naturally occurring mutation can provide potential sources for selection and propagation of beneficial variants that carry desired characters (GMITTER, 1994).

Chromosomal Changes due to Radiation

Genome stability is essential to support the survival and proper function of all organisms. DNA double-strand breaks (DSB) can cause chromosomal aberrations and affect many genes, resulting in genome instability, cell malfunctioning and cell death (VAN GENT et al., 2001). Both exogenous factors, such as radiation and chemical agents, and endogenous sources, such as free radicals, can impair the integrity of DNA, and the accumulation of DNA damage can eventually induce permanent damage to the cellular functioning and cell death (VAN GENT et al., 2001). Previous studies found that the DNA-damaging agents gave rise to chromosome rearrangements associated with loss of markers located near the chromosome ends, reflecting non-homologous end-joining or telomere addition to the broken end (MYUNG AND KOLODNER, 2003). There are two primary forms of genomic instability: mutational instability characterized by point mutations or small deletions, and chromosomal instability characterized by the gross rearrangement of chromosomes, including loss, gain or amplification of whole chromosomes or large chromosome fragments (VAN GENT et al., 2001). Translocation is another type of gross chromosomal aberration, where chromosome arms are exchanged (VAN GENT et al., 2001). The result of translocation can be the deregulation of gene expression or the fusion of two genes (VAN GENT et al., 2001).

Ionizing radiation has been utilized as a powerful tool for mutagenesis to explore genome organization (ARGUESO et al., 2008). The radiation track deposits discrete energy in biomolecules during its passage, which can cause chemical modifications to the biomolecules and defines the spatial distribution of lesions induced (LOMAX et al., 2013).

Also, the ionization of water molecules can produce hydroxyl radicals, which can attack the DNA and cause DNA damage indirectly (MAHANEY et al., 2009). For DNA, ionizing radiation will cause two or more lesions within one or two helical turns of the DNA and create significant levels of clustered DNA damage, which can be considered as a signature of ionizing radiation compared with endogenously induced lesions, which homogeneously distribute (LOMAX et al., 2013). Ionizing radiation can cause DNA lesions in several forms, including base damage in one or both DNA strands, combinations of base damage and single-strand breaks, simple DSBs, and DSBs with proximal damage (DATTA et al., 2005). Among all the lesions caused by ionizing radiation, double-strand breaks, which are breaks of the phosphodiester backbone of both strands of the DNA separated by approximately 10 base pairs with single-stranded overhangs of variable length, are principally responsible for radiation lethality and the most genotoxic, particularly those which occur during genome replication and segregation of duplicated chromosomes into daughter cells (DATTA et al., 2005; LOMAX et al., 2013; VAN GENT et al., 2001). Additional forms of DNA damage may also be present on the DNA surrounding the DSB, resulting in complex or clustered lesions (MAHANEY et al., 2009). Unrepaired DSBs lead to cell death, and misrepaired DSBs may result in chromosomal translocations and genomic instability (MAHANEY et al., 2009). DSBs can impede proper genome duplication, and further acentric chromosome fragments will not evenly partition between daughter cells if they are carried through mitosis (VAN GENT et al., 2001). The yield of DSB increases linearly with radiation dose and with the increase in DNA relaxation (LOMAX et al., 2013). In most naturally occurring DSBs and DSBs as the result of low-LET ionizing (linear energy transfer) radiation, the strand breaks

contain nucleotides consisting of phosphate or phosphoglycolate groups at the 3' ends (PASTWA et al., 2003). Studies showed that DSBs generated by low-LET radiations distribute randomly within the cell, while the distribution of DSBs produced by high-LET radiations was found to be nonrandom and determined by higher-order chromatin structure, resulting in regionally MDS (multiply damaged sites) (PRISE et al., 2001). Within genes, DNA that is close to the matrix attachment regions was found to be more resistant to breakage than the rest of the genome (PRISE et al., 2001). However, relatively low doses of ionizing radiation can contribute to chromosomal instability without causing the cell death (VAN GENT et al., 2001). Studies of site-specific DSBs showed that DSBs could give rise to chromosomal instability and induce chromosomal translocations (RICHARDSON AND JASIN, 2000). DSBs in unique DNA sequences are confined to undergo homologous recombination repair by using sister chromatids or homologous chromosomes, while repetitive DNA elements are the main source of DSB response and genome plasticity and most chromosome aberrations resulting from homologous recombination between retrotransposons are located at nonallelic sites (ARGUESO et al., 2008). Chromosome rearrangements were reported to frequently occur at or near repetitive elements (RICHARDSON AND JASIN, 2000). It is unlikely that a single DSB can significantly affect the viability or the recovery of recombinants, but a second DSB is demonstrated to increase the translocation frequency at least 2,000-fold and two DSBs are sufficient to cause chromosomal translocations (RICHARDSON AND JASIN, 2000).

However, the double strand break can be repaired by two mechanistically distinct pathways in the cell: homologous recombination (HR) and non-homologous end-joining (NHEJ).

Other pathways were also found, including single-strand annealing (SSA), microhomology-mediated end-joining (MMEJ), alternative or backup NHEJ (A-NHEJ or B-NHEJ), gene conversion (GC) and break-induced replication (BIR) (MEHTA AND HABER, 2014; MLADENOV AND ILIAKIS, 2011; SYMINGTON AND GAUTIER, 2011).

HR and NHEJ are both competitive and collaborative while the differences between HR and NHEJ are the requirement for a homologous template DNA, in which HR relies on the homologous template as source to copy lost DNA sequences and thus can ensure the fidelity of DSB repair while NHEJ can't because the sealing of the break does not rely on a template (MCVEY AND LEE, 2008; VAN GENT et al., 2001). They function at different efficiency depending on the stage of the cell cycle: HR is most efficient in the S and G2 phases when sister chromatids are available as repair templates, and NHEJ functions in the G1 phase without depending on homology to couple DNA ends when a sister chromatid is absent, but nucleotides adjacent to the DSB may be lost during the broken chromosome end joining (RICHARDSON AND JASIN, 2000; VAN GENT et al., 2001). These two pathways are not mutually exclusive and their relative contribution might differ depending on developmental stage, cell type, and DNA damage type (VAN GENT et al., 2001). The studies of gamma-radiation produced DSBs in yeast showed that most DSBs are efficiently repaired by HR using sister chromatids or homologous chromosomes as templates and only 2% of DSBs result in chromosome aberrations (ARGUESO et al., 2008). For DNA damaged by densely ionizing agents such as ionizing radiation, it is common that nucleotides around the DSB site are damaged in one or both strands and at one or both sides of the break. In these situations, NHEJ is used to repair the DSB. During this process, DSBs are detected

and the DNA ends are tethered and protected, then the damaged or non-ligatable groups of DNA ends are removed before ligation, making the NHEJ repair potentially error-prone (MAHANEY et al., 2009). Due to the structural and chemical complexity of DSB, the repairability reduces when compared with that of individual lesions, and therefore the NHEJ repair is inhibited as the structural complexity of DSBs increases (LOMAX et al., 2013; PASTWA et al., 2003; VAN GENT et al., 2001). The complex or heterogeneous end structures of DSBs produced by gamma rays, such as a high frequency of modifications of the oxidized base and abasic (AP) sites directly adjacent to the DSB ends, makes direct end joining refractory (LOMAX et al., 2013; PASTWA et al., 2003). Due to the difficulty of repair process increases with increases in the structural complexity of the break, gamma irradiation-induced DSBs were found to be rejoined much less efficiently than readily ligatable restriction enzyme-induced DSBs (PASTWA et al., 2003). Pastwa et al. performed the *in vitro* double-strand break repair assay using plasmid DNA with different DSB-damaging agents and found end joining varied with the type of agent and decreased as the cytotoxicity of the DSB-inducing agent increased (PASTWA et al., 2003). The secondary lesions proximal to DSB ends were found to inhibit NHEJ repair strongly, and the repair efficiency of DSBs produced by gamma rays was only one-fifteenth of the repair efficiency of DSBs produced by restriction enzymes (PASTWA et al., 2003). The consequences of DSB repair may include the generation of chromosome aberration breakpoints and altered gene dosage.

GC and BIR are two major mechanisms of HR, and HR also includes other mechanisms such as SSA and gene targeting (MEHTA AND HABER, 2014). GC is the nonreciprocal

transfer of a short fragment of DNA from a ‘donor’ sequence to an ‘acceptor’ sequence of high homology without requiring many normal DNA replication machinery components (CHEN et al., 2007). Usually, in DSB repair, GC mediates the transfer of DNA fragment from intact sequences to the region where excision occurs, where the homologous sequences can be on the same chromatid, sister chromatids, homologous chromosomes or different chromosomes (CHEN et al., 2007). BIR is a recombination-dependent replication when only one end of a broken chromosome shares homology with the template (MEHTA AND HABER, 2014). During the BIR process, the invaded strand forms a unidirectional replication folk and copies the whole sequence from the homologous site to the telomere, the consequence of which either appears similar to a non-reciprocal translocation if the template is a non-homologous chromosome or causes the loss of heterozygosity if the template is a homologous chromosome (MEHTA AND HABER, 2014). SSA occurs when homologous repeats flanking the breakpoint can anneal, the repair of which leads to a deletion between the two homologous repeats (MEHTA AND HABER, 2014). In MMEJ DNA repair, 5–25 bp microhomologous sequences are used to align the broken ends before joining, which results in deletions flanking the original break (MCVEY AND LEE, 2008). MMEJ is also associated with chromosome abnormalities such as deletions, translocations, inversions and other complex rearrangements (MCVEY AND LEE, 2008). A-NHEJ (sometimes referred to as B-NHEJ) pathway is the microhomology-mediated pathway and functions as a backup and when the classical NHEJ pathway is chemically or genetically compromised (ILIAKIS, 2009; NUSSENZWEIG AND NUSSENZWEIG, 2007). However, the efficiency of the A-NHEJ is 20-fold less than the classical NHEJ and the repair junctions

display deletions, insertions, and overlapping microhomologies (DECOTTIGNIES, 2013; NUSSENZWEIG AND NUSSENZWEIG, 2007).

Physical mutagenic agents and aberrant repair of lesions usually cause chromosome rearrangements, which is one of the most frequent mutations (RICHARDSON AND JASIN, 2000; VICCINI AND DE CARVALHO, 2002). Fast neutrons, X-rays and gamma rays are the most common and efficient methods to inducing mutations to fruit trees (VARDI et al., 2008). Natarajan found that induced mutations could cause chromosomal aberrations in somatic cells in root meristems and pollen gametic cells (NATARAJAN, 2005). It has long been known that radiation can induce chromosomal mutations such as translocations, inversions and large deletions (SAX, 1938). Also, the chromosomal mutation rate was observed to increase with the increase of radiation dosage (SAX, 1938). If the mutation occurred in cells that give rise to gametes, the resulting gametes could also carry the mutation and lead to hereditary changes. Chromosomal aberrations can also reduce fertility because the mutation usually affects one of a pair of chromosomes and results in heterozygous chromosome rearrangements, which brings about abnormal chromosome pairing during meiosis, and the gametes carrying mutations will be sterile (DA COSTA SILVA et al., 2011). The joining of such a genetically unbalanced gamete with a normal haploid gamete will form a genetically unbalanced zygote that is often subsequently aborted (HADORN AND MITTWOCH, 1961).

Objectives of Study

The objectives mainly focus on detection of genomic differences between W. Murcott mandarin and its mutant derivative, Tango mandarin. The appearances of W. Murcott and Tango mandarins are almost identical and their fruits are indistinguishable under isolated growth condition without cross-pollination. Due to their premium quality, these two cultivars are popular and broadly planted throughout California. The most crucial difference between W. Murcott and Tango reported so far is the seed count under the cross-pollination condition, where W. Murcott mandarin averagely bears 11-22 seeds per fruit while Tango mandarin fruit bears less than one (ROOSE AND WILLIAMS, 2007a). Another difference, which is also an essential commercial attribute, is that the viability of Tango mandarin pollen is too low to cause other citrus to produce seeds. The pollen germination rate, pollen size and pollen grain count per anther of Tango mandarin were reported remarkably reduced compared to W. Murcott mandarin (ROOSE AND WILLIAMS, 2007a). Sterility of both ovule and pollen in Tango possibly indicate the generation and development of both female and male gametes have been affected by gamma irradiation mutation.

With assistance of NGS data and Axiom Citrus 15AX SNP array data we currently have, to identify molecular differences between W. Murcott mandarin and Tango mandarin and discover potential causes of Tango gamete sterility that associate with them, the following aspects are to be investigated, including but not limited to: 1. discovery of small genomic differences such as SNPs and Indels and their predicted effects on affected genes, 2. detection and characterization of large genomic differences such as copy number variation

and chromosome rearrangements, and 3. Determination of whether Tango mandarin is a chimera with different genetic composition in different cell layers. In addition, we need to develop fast and efficient methods to distinguish Tango from W. Murcott for patent protection.

References

Adam-Blondon, A.F., F. Lahogue-Esnault, A. Bouquet, J.M. Boursiquot, and P. This, 2001. Usefulness of two SCAR markers for marker-assisted selection of seedless grapevine cultivars. *Vitis* 40:147-155.

Argueso, J.L., J. Westmoreland, P.A. Mieczkowski, M. Gawel, T.D. Petes, and M.A. Resnick, 2008. Double-strand breaks associated with repetitive DNA can reshape the genome. *Proceedings of the National Academy of Sciences of the United States of America* 105:11845-11850.

Aydin Uzun, O.G., T. Yesilogiu, Y. Aka-Kacar, and O. Tuzcu, 2010. Distinguishing grapefruit and pummelo accessions using ISSR markers. *Czech J. Genet. Plant Breed* 46:170-177.

Barrett, H. and A. Rhodes, 1976. A numerical taxonomic study of affinity relationships in cultivated Citrus and its close relatives. *Systematic Botany*:105-136.

Bayer, R.J., D.J. Mabberley, C. Morton, C.H. Miller, I.K. Sharma, B.E. Pfeil, S. Rich, R. Hitchcock, and S. Sykes, 2009. A molecular phylogeny of the orange subfamily (Rutaceae: Aurantioideae) using nine cpDNA sequences. *American Journal of Botany* 96:668-685.

Beattie, G., P. Holford, D. Mabberley, A. Haigh, and P. Broadbent, 2008. On the origins of Citrus, Huanglongbing, Diaphorina citri and Trioza erytreae, p. 23-56, Proceedings of the International Research Conference on Huanglongbing.

Bender, R.J., R.P. dos Santos, D. Guerra, S.F. Schwarz, and S.D. Bender, 2017. 'URS Campestre' seedless orange: a new mutant with female sterility. *Scientia Agricola* 74:371-377.

Bermejo, A., J. Pardo, and A. Cano, 2011. Influence of gamma irradiation on seedless citrus production: pollen germination and fruit quality. *Food and Nutrition Sciences* 2:169-180.

Bouquet, A. and Y. Danglot, 1996. Inheritance of seedlessness in grapevine (Vitis vinifera L). *Vitis* 35:35-42.

Bowman, K.D., F.G. Gmitter, G.A. Moore, and R.L. Rouseff, 1991. Citrus fruit sector chimeras as a genetic resource for cultivar improvement. *Journal of the American Society for Horticultural Science* 116:888-893.

Burger, P., A. Bouquet, and M.J. Striem, 2009. Grape breeding, p. 161-189, Breeding plantation tree crops: Tropical species. Springer.

Cai, X.D., J. Fu, X.X. Deng, and W.W. Guo, 2007. Production and molecular characterization of potential seedless cybrid plants between pollen sterile Satsuma mandarin and two seedy Citrus cultivars. *Plant Cell Tissue and Organ Culture* 90:275-283.

Carmi, N., Y. Salts, B. Dedicova, S. Shabtai, and R. Barg, 2003. Induction of parthenocarpy in tomato via specific expression of the rolB gene in the ovary. *Planta* 217:726-735.

Carvalho, R., W.S. Soares, A.C. Brasileiro-Vidal, and M. Guerra, 2005. The relationships among lemons, limes and citron: a chromosomal comparison. *Cytogenetic and Genome Research* 109:276-282.

Chai, L., X. Ge, M.K. Biswas, Q. Xu, and X. Deng, 2011. Self-sterility in the mutant 'Zigui shatian'pummelo (Citrus grandis Osbeck) is due to abnormal post-zygotic embryo development and not self-incompatibility. *Plant Cell, Tissue and Organ Culture (PCTOC)* 104:1-11.

Chen, J.-M., D.N. Cooper, N. Chuzhanova, C. Férec, and G.P. Patrinos, 2007. Gene conversion: mechanisms, evolution and human disease. *Nature Reviews Genetics* 8:762.

Crowley, J.R., 2011. A Molecular Genetic Approach to Evaluate a Novel Seedless Phenotype Found in Tango, a New Variety of Mandarin Developed From Gamma-Irradiated W. Murcott.

Cuenca, J., Y. Froelicher, P. Aleza, J. Juarez, L. Navarro, and P. Ollitrault, 2011. Multilocus half-tetrad analysis and centromere mapping in citrus: evidence of SDR mechanism for 2n megagametophyte production and partial chiasma interference in mandarin cv 'Fortune'. *Heredity* 107:462-470.

Curk, F., G. Ancillo, F. Ollitrault, X. Perrier, J.P. Jacquemoud-Collet, A. Garcia-Lor, L. Navarro, and P. Ollitrault, 2015. Nuclear species-diagnostic SNP markers mined from 454 amplicon sequencing reveal admixture genomic structure of modern citrus varieties. *PLoS One* 10:e0125628.

Curk, F., F. Ollitrault, A. Garcia-Lor, F. Luro, L. Navarro, and P. Ollitrault, 2016. Phylogenetic origin of limes and lemons revealed by cytoplasmic and nuclear markers. *Ann Bot* 117:565-583.

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

Datta, K., R.D. Neumann, and T.A. Winters, 2005. Characterization of complex apurinic/apyrimidinic-site clustering associated with an authentic site-specific

radiation-induced DNA double-strand break. *Proc Natl Acad Sci U S A* 102:10569-10574.

Decottignies, A., 2013. Alternative end-joining mechanisms: a historical perspective. *Front Genet* 4:48.

Deng, X., W. Guo, and X. Sun, 1996. Advances in breeding and selection of seedless types of Citrus in China. *Acta Horticulturae Sinica (China)*.

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

Ensminger, M.E. and A.H. Ensminger, 1993. Foods & Nutrition Encyclopedia, Two Volume Set. CRC press.

Esen, A. and R.K. Soost, 1971. Unexpected triploids in Citrus: their origin, identification, and possible use. *Journal of Heredity* 62:329-333.

Esen, A. and R.K. Soost, 1972. Aneuploidy in Citrus. *American Journal of Botany*:473-477.

Esen, A. and R.K. Soost, 1973. Seed development in Citrus with special reference to 2x x 4x crosses. *American Journal of Botany*:448-462.

Esen, A., R.K. Soost, and G. Geraci, 1978. Seed set, size, and development after 4x × 2x and 4x × 4x crosses in Citrus. *Euphytica* 27:283-294.

Federici, C.T., D.Q. Fang, R.W. Scora, and M.L. Roose, 1998. Phylogenetic relationships within the genus Citrus (Rutaceae) and related genera as revealed by RFLP and RAPD analysis. *Theoretical and Applied Genetics* 96:812-822.

Ficcadenti, N., S. Sestili, T. Pandolfini, C. Cirillo, G.L. Rotino, and A. Spena, 1999. Genetic engineering of parthenocarpic fruit development in tomato. *Molecular Breeding* 5:463-470.

Frost, H. and C. Krug, 1942. Diploid-tetraploid periclinal chimeras as bud variants in citrus. *Genetics* 27:619-634.

Gambetta, G., A. Gravina, C. Fasiolo, C. Fornero, S. Galiger, C. Inzaurralde, and F. Rey, 2013. Self-incompatibility, parthenocarpy and reduction of seed presence in 'Afourer' mandarin. *Scientia Horticulturae* 164:183-188.

Garcia-Lor, A., F. Curk, H. Snoussi-Trifa, R. Morillon, G. Ancillo, F. Luro, L. Navarro, and P. Ollitrault, 2013. A nuclear phylogenetic analysis: SNPs, indels and SSRs deliver new insights into the relationships in the 'true citrus fruit trees' group (Citrinae, Rutaceae) and the origin of cultivated species. *Annals of Botany* 111:1-19.

Garcia-Lor, A., F. Luro, L. Navarro, and P. Ollitrault, 2012. Comparative use of InDel and SSR markers in deciphering the interspecific structure of cultivated citrus genetic diversity: a perspective for genetic association studies. *Molecular Genetics and Genomics* 287:77-94.

Gidoni, D. and N. Carmi, 2007. Mutagenesis for seedlessness in Citrus. *Israel Journal of Plant Sciences* 55:133-135.

Gmitter, F.G., 1994. Contemporary approaches to improving citrus cultivars. *HortTechnology* 4:206-210.

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

Gorguet, B., A.W. van Heusden, and P. Lindhout, 2005. Parthenocarpic fruit development in tomato. *Plant Biol (Stuttg)* 7:131-139.

Guo, W.W., X.X. Deng, and H.L. Yi, 2000. Somatic hybrids between navel orange (Citrus sinensis) and grapefruit (C-paradisi) for seedless triploid breeding. *Euphytica* 116:281-285.

Hadorn, E. and U. Mittwoch, 1961. Developmental genetics and lethal factors. *The American Journal of the Medical Sciences* 242:522.

Hearn, C., 1984. Development of seedless orange and grapefruit cultivars through seed irradiation. *Journal American Society for Horticultural Science* 109:270-273.

Hearn, C., 1986. Development of seedless grapefruit cultivars through budwood irradiation. *Journal of the American Society for Horticultural Science (USA).*

Hensz, R., 1991. Mutation breeding of grapefruit (Citrus paradisi Macf.), Plant mutation breeding for crop improvement. V. 1.

Hörandl, E. and O. Paun, 2007. Patterns and sources of genetic diversity in apomictic plants: implications for evolutionary potentials. *Apomixis: Evolution, mechanisms and perspectives. ARG Gantner Verlag KG, Lichtenstein*:169-194.

Hu, Z.Y., M. Zhang, Q.G. Wen, J. Wei, H.L. Yi, X.X. Deng, and X.H. Xu, 2007. Abnormal microspore development leads to pollen abortion in a seedless mutant of 'Ougan' mandarin (Citrus suavissima hort. ex tanaka). *Journal of the American Society for Horticultural Science* 132:777-782.

Iliakis, G., 2009. Backup pathways of NHEJ in cells of higher eukaryotes: cell cycle dependence. *Radiother Oncol* 92:310-315.

Jacobsen, S.E. and N.E. Olszewski, 1993. Mutations at the spindly locus of Arabidopsis alter gibberellin signal-transduction. *Plant Cell* 5:887-896.

Khalil, S.A., A. Sattar, and R. Zamir, 2011. Development of sparse-seeded mutant kinnow (Citrus reticulata Blanco) through budwood irradiation. *African Journal of Biotechnology* 10:14562-14565.

Khawale, R.N., S.K. Singh, and Y. Vimala, 2004. Gamma rays induced in vitro mutagenesis and molecular marker-assisted selection of mutants in grapevine, p. 643-652, V International Symposium on In Vitro Culture and Horticultural Breeding 725.

Koltunow, A.M., T. Hidaka, and S.P. Robinson, 1996. Polyembryony in citrus (accumulation of seed storage proteins in seeds and in embryos cultured in vitro). *Plant Physiology* 110:599-609.

Lahogue, F., P. This, and A. Bouquet, 1998. Identification of a codominant scar marker linked to the seedlessness character in grapevine. *TAG Theoretical and Applied Genetics* 97:950-959.

Li, W., G. Liu, and S. He, 1992. Leaf isozymes of mandarin, p. 217-220, Proc Int Soc Citricult.

Liang, G.L., G.M. Xiong, Q.G. Guo, Q. He, and X.L. Li, 2007. AFLP analysis and the taxonomy of Citrus, p. 137-142. In: K.E. Hummer (ed.), Proceedings of the Second International Symposium on Plant Genetic Resources of Horticultural Crops, Vols 1 and 2. International Society Horticultural Science, Leuven 1.

Liu, Y.Q., E. Heying, and S.A. Tanumihardjo, 2012. History, global distribution, and nutritional importance of citrus fruits. *Comprehensive Reviews in Food Science and Food Safety* 11:530-545.

Lomax, M.E., L.K. Folkes, and P. O'Neill, 2013. Biological consequences of radiation-induced DNA damage: relevance to radiotherapy. *Clin Oncol (R Coll Radiol)* 25:578-585.

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

Mabberley, D.J., 1998. Australian Citreae with notes on other Aurantioideae (Rutaceae). *Telopea* 7:333-344.

Mabberley, D.J., 2004. Citrus (Rutaceae): A review of recent advances in etymology, systematics and medical applications. *Blumea* 49:481-498.

Mahaney, B.L., K. Meek, and S.P. Lees-Miller, 2009. Repair of ionizing radiation-induced DNA double-strand breaks by non-homologous end-joining. *Biochem J* 417:639-650.

Marcovitch, V.V., 1926. Indeeling van het geslacht Citrus. *Landbouw* 2(4).

McVey, M. and S.E. Lee, 2008. MMEJ repair of double-strand breaks (director's cut): deleted sequences and alternative endings. *Trends in Genetics* 24:529-538.

Mehta, A. and J.E. Haber, 2014. Sources of DNA double-strand breaks and models of recombinational DNA repair. *Cold Spring Harb Perspect Biol* 6:a016428.

Mladenov, E. and G. Iliakis, 2011. Induction and repair of DNA double strand breaks: the increasing spectrum of non-homologous end joining pathways. *Mutat Res* 711:61-72.

Moore, G.A., 2001. Oranges and lemons: clues to the taxonomy of Citrus from molecular markers. *TRENDS in Genetics* 17:536-540.

Myung, K. and R.D. Kolodner, 2003. Induction of genome instability by DNA damage in Saccharomyces cerevisiae. *DNA Repair (Amst)* 2:243-258.

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

Nicolosi, E., 2007. Origin and Taxonomy, p. 19-43, Citrus genetics, breeding and biotechnology.

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

Nussenzweig, A. and M.C. Nussenzweig, 2007. A backup DNA repair pathway moves to the forefront. *Cell* 131:223-225.

Okie, W., 2000. Register of new fruit and nut varieties. List 40. *HortScience* 35:812-826.

Ollitrault, P., Y. Froelicher, D. Dambier, F. Luro, and M. Yamamoto, 2007. Seedlessness and ploidy manipulation. *Citrus Genetics, Breeding and Biotechnology*:197-218.

Ollitrault, P., C. Jacquemond, C. Dubois, and F. Luro, 2003. Botany and genetic resources. *Genetic Diversity of Cultivated Tropical Plants*:193.

Ollitrault, P., J. Terol, A. Garcia-Lor, A. Berard, A. Chauveau, Y. Froelicher, C. Belzile, R. Morillon, L. Navarro, D. Brunel, and M. Talon, 2012. SNP mining in C. clementina BAC end sequences; transferability in the Citrus genus (Rutaceae), phylogenetic inferences and perspectives for genetic mapping. *BMC Genomics* 13:13.

Pandolfini, T., 2009. Seedless fruit production by hormonal regulation of fruit set. *Nutrients* 1:168-177.

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

Pastwa, E., R.D. Neumann, K. Mezhevaya, and T.A. Winters, 2003. Repair of radiation-induced DNA double-strand breaks is dependent upon radiation quality and the structural complexity of double-strand breaks. *Radiation Research* 159:251-261.

Perl, A., N. Sahar, P. Spiegel-Roy, S. Gavish, R. Elyasi, E. Orr, and H. Bazak, 1998. Conventional and biotechnological approaches in breeding seedless table grapes, p. 613-618, VII International Symposium on Grapevine Genetics and Breeding 528.

Prise, K.M., M. Pinto, H.C. Newman, and B.D. Michael, 2001. A review of studies of ionizing radiation-induced double-strand break clustering. *Radiat Res* 156:572-576.

Qin, Y., C. Xu, Z. Ye, J.A.T. Da Silva, and G. Hu, 2015. Seedless mechanism of a new citrus cultivar 'Huami wuhegonggan'(Citrus sinensis× C. reticulata). *Pak. J. Bot* 47:2369-2378.

Richardson, C. and M. Jasin, 2000. Frequent chromosomal translocations induced by DNA double-strand breaks. *Nature* 405:697-700.

Roose, M.L. and T.E. Williams, 2007a. Mandarin tree named 'Tango'. U.S. Patent No. PP17,863.

Roose, M.L. and T.E. Williams, 2007b. Mutation breeding. *Citrus Genetics, Breeding and Biotechnology*:345-352.

Saunt, J. 1990. Citrus varieties of the world: an ilustrated guide. Norwich, GB: Sinclair International.

Sawamura, M. and T. Kuriyama, 1988. Quantitative-determination of volatile constituents in the pummelo (Citrus grandis Osbeck forma Tosa-buntan). *Journal of Agricultural and Food Chemistry* 36:567-569.

Sax, K., 1938. Chromosome aberrations induced by X-Rays. *Genetics* 23:494-516.

Scora, R., J. Kumamoto, R. Soost, and E. Nauer, 1982. Contribution to the origin of the grapefruit, Citrus paradisi (Rutaceae). *Systematic Botany*:170-177.

Scora, R.W., 1975. On the history and origin of Citrus. *Bulletin of the Torrey Botanical Club*:369-375.

Shamel, A.D., C.S. Pomeroy, and R.E. Caryl, 1925. Bud selection in the Washington Navel orange. Progeny tests of the Thomson, Washington, Corrugated and Golden Buckeye strains originating as limb variations on a single tree. *Journal of Heredity* 16:233-241.

Soost, R. and J. Cameron, 1985. Melogold, a triploid pummelo—grapefruit hybrid. *HortScience* 20:1134-1135.

Soost, R.K. and J.W. Cameron, 1980. 'Oroblanco', a triploid pummelo-grapefruit hybrid. *HortScience* 15:667-669.

Spiegel-Roy, P., 1990. Economic and agricultural impact of mutation breeding in fruit trees. *Mutation Breeding Review*:215-235.

Spiegel-Roy, P. and E.E. Goldschmidt, 1996. The biology of citrus. Cambridge University Press.

Spiegel-Roy, P. and A. Vardi. 1990. Induced mutations in Citrus, International Atomic Energy Agency (IAEA).

Swingle, W.T., 1967. The botany of Citrus and its wild relatives. *The citrus industry*:190-430.

Symington, L.S. and J. Gautier, 2011. Double-strand break end resection and repair pathway choice. *Annu Rev Genet* 45:247-271.

Talon, M., L. Zacarias, and E. Primo-Millo, 1992. Gibberellins and parthenocarpic ability in developing ovaries of seedless mandarins. *Plant Physiol* 99:1575-1581.

Tanaka, C. and T. Tanaka, 1954. Species problem in Citrus: A critical study of wild and cultivated units of citrus, based upon field studies in their native homes. Japanese Society for the Promotion of Science.

Tanaka, T., 1961. Contribution to the knowledge of Citrus classification. *Reports Citrologia*:107-114.

van Gent, D.C., J.H. Hoeijmakers, and R. Kanaar, 2001. Chromosomal stability and the DNA double-stranded break connection. *Nat Rev Genet* 2:196-206.

Vardi, A., A. Elhanati, A. Frydman-Shani, and H. Neumann, 1996. Strategies and consideration in mandarin improvement programmes, p. 109-112, Proceedings of International Society of Citriculture.

Vardi, A., A. Elhanati, S. Frydman, H. Neumann, and R. Spiegel, 1995. New considerations on the choice of irradiation dose rate in Citrus, Induced mutations and molecular techniques for crop improvement. Proceedings of an international symposium.

Vardi, A., I. Levin, and N. Carmi, 2008. Induction of seedlessness in citrus: From classical techniques to emerging biotechnological approaches. *Journal of the American Society for Horticultural Science* 133:117-126.

Vardi, A., P. Spiegel-roy, and A. Elchanati, 1993. Mandarin tree names Mor. US Patent PP8,378.

Varoquaux, F., R. Blanvillain, M. Delseny, and P. Gallois, 2000. Less is better: new approaches for seedless fruit production. *Trends in Biotechnology* 18:233-242.

Viccini, L.F. and C.R. De Carvalho, 2002. Meiotic chromosomal variation resulting from irradiation of pollen in maize. *Journal of Applied Genetics* 43:463-469.

Webber, H.J., 1943. Cultivated varieties of Citrus. *The citrus industry* 1:475-668.

Webber, H.J. and L.D. Batchelor, 1943. History, botany and breeding. *The citrus industry* 1.

Wu, G.A., S. Prochnik, J. Jenkins, J. Salse, U. Hellsten, F. Murat, X. Perrier, M. Ruiz, S. Scalabrin, J. Terol, M.A. Takita, K. Labadie, J. Poulain, A. Couloux, K. Jabbari, F. Cattonaro, C. Del Fabbro, S. Pinosio, A. Zuccolo, J. Chapman, J. Grimwood, F.R. Tadeo, L.H. Estornell, J.V. Munoz-Sanz, V. Ibanez, A. Herrero-Ortega, P. Aleza, J. Perez-Perez, D. Ramon, D. Brunel, F. Luro, C. Chen, W.G. Farmerie, B. Desany, C. Kodira, M. Mohiuddin, T. Harkins, K. Fredrikson, P. Burns, A. Lomsadze, M. Borodovsky, G. Reforgiato, J. Freitas-Astua, F. Quetier, L. Navarro, M. Roose, P. Wincker, J. Schmutz, M. Morgante, M.A. Machado, M. Talon, O. Jaillon, P. Ollitrault, F. Gmitter, and D. Rokhsar, 2014. Sequencing of diverse mandarin, pummelo and orange genomes reveals complex history of admixture during citrus domestication. *Nature Biotechnology* 32:656-662.

Wu, G.A., J. Terol, V. Ibanez, A. Lopez-Garcia, E. Perez-Roman, C. Borreda, C. Domingo, F.R. Tadeo, J. Carbonell-Caballero, R. Alonso, F. Curk, D.L. Du, P. Ollitrault, M.L. Roose, J. Dopazo, F.G. Gmitter, D.S. Rokhsar, and M. Talon, 2018. Genomics of the origin and evolution of Citrus. *Nature* 554:311-+.

Xiao, J., J. Tan, H. Liu, L. Chen, W. Ye, and W. Cheng, 2007. Studies on the seedless mechanism of Lipeng No. 2 Ponkan (Citrus reticulata). *Journal of Fruit Science* 24:421-426.

Xu, Q., L.L. Chen, X. Ruan, D. Chen, A. Zhu, C. Chen, D. Bertrand, W.B. Jiao, B.H. Hao, M.P. Lyon, J. Chen, S. Gao, F. Xing, H. Lan, J.W. Chang, X. Ge, Y. Lei, Q. Hu, Y. Miao, L. Wang, S. Xiao, M.K. Biswas, W. Zeng, F. Guo, H. Cao, X. Yang, X.W. Xu, Y.J. Cheng, J. Xu, J.H. Liu, O.J. Luo, Z. Tang, W.W. Guo, H. Kuang, H.Y. Zhang, M.L. Roose, N. Nagarajan, X.X. Deng, and Y. Ruan, 2013. The draft genome of sweet orange (Citrus sinensis). *Nature Genetics* 45:59-66.

Yamamoto, M., R. Matsumoto, N. Okudai, and Y. Yamada, 1997. Aborted anthers of Citrus result from gene-cytoplasmic male sterility. *Scientia Horticulturae* 70:9-14.

Ye, W.J., Y.H. Qin, Z.X. Ye, J.A.T. da Silva, L.X. Zhang, X.Y. Wu, S.Q. Lin, and G.B. Hu, 2009. Seedless mechanism of a new mandarin cultivar 'Wuzishatangju' (Citrus reticulata Blanco). *Plant Science* 177:19-27.

Zhang, S.W., G.X. Huang, F. Ding, X.H. He, and J.C. Pan, 2012. Mechanism of seedlessness in a new lemon cultivar 'Xiangshui' [Citrus limon (L.) Burm. F.]. *Sexual Plant Reproduction* 25:337-345.

Chapter 2: DNA Polymorphisms Between W. Murcott and Tango

Abstract

Since Tango mandarin was developed by gamma-irradiation mutation breeding of W. Murcott mandarin, they are quite similar in appearance, and both are easily peelable mid-late season mandarins with deep orange rind color and sweet juicy fruit. The most significant phenotypic difference is the condition of bearing mostly seedless fruits, for which W. Murcott mandarin requires an isolated condition without cross-pollination while Tango mandarin bears very low-seeded fruit naturally without isolation. Also, in contrast to W. Murcott mandarin, the viability of Tango pollen is too low to cause other seed-bearing citrus to produce seedy fruit. Since seedless citrus fruit have become attractive to consumers and commercially valuable, it is crucial to develop a fast, sensitive, reliable and economical method to distinguish the Tango variety from W. Murcott which will enable better enforcement of the Tango patent. Illumina NGS data was used to predict small-scale genomic differences between Tango and W. Murcott by Pindel program and Dindel programs. Then the predicted results were verified by allele-specific PCR, high-resolution melting real-time PCR and conventional PCR followed by sequencing. Three SNPs and seven Indels were confirmed as heterozygous in Tango while homozygous in W. Murcott, and the differences between Tango and W. Murcott at these loci could be clearly discerned as differences in size of PCR products on agarose gels or by real-time PCR melt curves.

These discovered SNPs and Indels provide an effective molecular marker system to distinguish Tango trees from W. Murcott trees.

Introduction

Utilization of trait-linked markers is a significant procedure for plant breeding programs. The use of markers for breeding can be traced back to the nineteenth century when morphological markers were used by Mendel in his experiments (AGARWAL et al., 2008). Later in the early 20th century, genetic linkage theory was established in Drosophila experiments (AGARWAL et al., 2008). A genetic marker is an inherited character that can be observed or a phenotype that can be assayed, for which the segregation of alleles at specific loci follow the Mendelian expectation (WIJERATHNA, 2015). The most commonly used genetic markers can be categorized into two types, morphological markers and molecular markers, both of which are demonstrated to be linked with the desired characteristic that is typically not as easily scored (MCCASKILL AND GIOVANNONI, 2002). Morphological markers are phenotypes that can be visually identified while molecular markers are not based on morphologically observable phenotype but based on a particular DNA fragment variation located at a certain chromosomal position and visualized by biochemical methods (MCCASKILL AND GIOVANNONI, 2002). A molecular marker is independent of phenotypic traits or gene expression, representing the differences between varieties at the genome level (AGARWAL et al., 2008; RUIZ et al., 2000). Since simply inherited morphological markers are relatively rare in non-model organisms and thus difficult for general and direct use, molecular markers were developed as an alternative to morphological markers. Compared to conventional morphological markers, molecular markers have many advantages, including but not limited to stability and detectability in

all tissues and growth and development stages (AGARWAL et al., 2008). Molecular markers enable genome-wide studies on organisms and wild populations, such as identification of recombination breakpoints for linkage mapping or quantitative trait locus (QTL) mapping, locating differentiated genomic regions between populations, or phylogeography of wild populations (DAVEY et al., 2011). The earliest molecular markers used were proteins such as isozymes, but these were eventually replaced by DNA-based markers which can be widely used without limitation (MCCASKILL AND GIOVANNONI, 2002). The first DNA marker system developed was RFLP, a method based on restriction digestion and hybridization of clones probes (SINGH AND SINGH, 2015).

To create high-resolution genetic maps and study genetic linkage of crop traits, various types of molecular marker systems have been developed (EDWARDS et al., 1992; PATERSON et al., 1988). It can be a chromosomal landmark or allele that allows the tracing of a specific region or a specific piece of DNA with a known position (SEMAGN et al., 2006). An ideal molecular marker should satisfy the following criteria: (1) be highly polymorphic and evenly and frequently distributed throughout the genome, (2) adequately resolve genetic differences between individuals, (3) generate multiple, independent, reliable and highly reproducible markers, (4) show codominant inheritance, distinguishing homozygous and heterozygous states of diploid organisms, (5) be neutral to selection over a range of environmental conditions, (6) be easy, fast and inexpensive to detect, (7) require small amounts of DNA, (8) link to distinct phenotypes, (9) require no prior information about the genome of an organism, and (10) facilitate easy exchange of data between laboratories (AGARWAL et al., 2008). However, no molecular marker technique can meet every

requirement. In general, there are two different types of markers: hybridization-based markers (non-PCR) and PCR based markers. Non-PCR based markers are visualized by hybridizing DNA to a labeled probe, which is a known sequence or DNA fragment (KUMAR et al., 2009), including restriction fragment length polymorphism (RFLP), restriction-site-associated DNA (RAD), variable number of tandem repeats (VNTRs) and array-based technology such as diversity array technology (DArT) and single feature polymorphism (SFP), etc (SINGH AND SINGH, 2015). Since molecular markers such as isozyme loci are limited and the fingerprint probe, which detects the complex genotype-specific hybridization pattern of a multicopy sequence by Southern blots, is expensive, time-consuming and requires large amounts of relatively pure genomic DNAs, PCR based technique has been utilized to accelerate the development of new DNA marker systems (RUIZ et al., 2000). PCR-based markers use specifically or arbitrarily chosen primers to amplify particular DNA sequences or loci in vitro, including random amplified polymorphic DNA (RAPD), simple sequence repeats (SSR) and amplified fragment length polymorphism (AFLP), etc (KUMAR et al., 2009). Both non-PCR based and PCR based molecular markers are widely used in plant taxonomical, evolutionary and phylogenic studies. The utilization of molecular markers has significantly improved knowledge of DNA polymorphism in plant biotechnology and plant genetics. Prior to about 2003, there was no ideal molecular approach to discover enough polymorphic markers to cover the whole genome in a single, simple and reliable experiment (DAVEY et al., 2011). The traditional approaches for the development of molecular markers were expensive and time-consuming, and could not easily be parallelized (DAVEY et al., 2011). However, the

development of next-generation sequencing (NGS) allows the discovery, sequencing and genotyping of thousands of markers in a single step in any genome of interest with mostly parallelized libraries, including those of species whose sequence information is unavailable (DAVEY et al., 2011). The high abundance and distribution throughout the genome of the polymorphisms permit the development of markers for many genes simultaneously (HAYASHI et al., 2006). Multiple markers together can be used to develop a multiplex detection system to improve the efficiency of genotyping multilines, as well as to design PCR-based markers for discrimination of each line from a group (HAYASHI et al., 2006). In molecular markers using NGS approach, the genomic DNA is digested with a restriction enzyme and only the ends of the fragments are sequenced in the simplest form, producing partial but genome-wide coverage (DAVEY et al., 2011). The reads from the reduced-representation sequencing are then mapped to the reference genome if a reference genome of high quality is available or assembled de novo in the absence of a reference genome, and finally, potential molecular markers and structural variations can be called (DAVEY et al., 2011). These reduced-representation libraries (RRLs) were firstly used for generation of a single nucleotide polymorphism (SNP) map of the human genome and later used to generate millions of candidate SNPs in different organisms (DAVEY et al., 2011). SNP is a single base change, usually between two alternative nucleotides, at a certain position in the DNA sequence. Also, the optimal coverage varies widely depending on experimental goals and strategies (DAVEY et al., 2011). When a high-quality genome is available, at least 30x coverage is desired for heterozygous samples if the whole genome needs to be covered with full genotyping. Lower coverage will cause lower confidence in calling genotypes or

discovery of new marker, and thus increasing the difficulty and cost of downstream analysis (DAVEY et al., 2011).

For grape, two AFLP molecular markers were successfully used for discrimination of the early bud bursting and fruiting somatic mutants from their parent, Flame Seedless grape, at intra-cultivar level (SCOTT et al., 2000). The successful development of the AFLP markers overcame the difficulties of consistent demarcation of somatic mutations and provided commercial identification and protection of the early bud-bursting Flame Seedless mutant (SCOTT et al., 2000). GSLP1, a SCAR maker was found linked to grape seedlessness and utilized for seedless grape progeny selection (LI et al., 2015). Another PCR-based molecular marker, a sequence-related amplified polymorphism (SRAP) was coupled with cDNA synthesis to determine the expression of an ovule abortion-related gene at the transcriptional level of *V. vinifera* L. cv. Youngle seedless grape (WANG et al., 2016). Also, multiple molecular markers were employed for marker-assisted selection of seedless grape progeny, such as hybrids of seeded parent Muscat of Hamburg × seedless parent Sultani (AKKURT et al., 2013).

Molecular markers and high-throughput genome sequencing have greatly helped to characterize genetic diversity in the germplasm pool for crop species. In maize, genetic diversity before and after domestication from its wild ancestor teosinte and among geographically distributed landraces have already been measured by molecular marker alleles and nucleotide sequence variations (MOOSE AND MUMM, 2008). Investigation of plant evolution and comparative genomics, as well as plant population structure and genetic responses to selection, are facilitated by information of molecular markers (MOOSE AND

MUMM, 2008). Since markers or genes that are close together on the same chromosome tend to stay together when each plant generation is produced, a genetic linkage map can be created specifying the relative chromosomal position of and distance between markers or genes and revealing associations between genes and economically important traits and quantitative trait loci, which ultimately enables marker-assisted selection (SEMAGN et al., 2006).

In citrus breeding, some characters of horticultural interest were found to be linked to molecular markers, including dwarfing, fruit acidity, *Poncirus trifoliata* CTV resistance, nucellar embryony and anthocyanin content, significantly improving marker-assisted selection (MAS) (BERNET et al., 2004; BUTELLI et al., 2012; CHENG AND ROOSE, 1995; FANG et al., 1997; KEPIRO AND ROOSE, 2010). In the studies of Ollitrault and Navarro, the locus of a candidate alternaria brown spot resistance gene ABSr was found to be flanked by an SSR marker (TTC8) and a SNP marker (CiC3248-06) (CUENCA et al., 2013). In later studies, a closer SNP marker was found without recombination between them using fine mapping of useful SNPs in the previously defined flanking region of the ABSr locus, a result that is useful for discriminating resistant cultivars from susceptible cultivars (CUENCA et al., 2016). In citrus, molecular markers are also extensively used to determine the genetic origins of plants and distinguish between zygotic and nucellar seedlings to eliminate unwanted genotypes, because all hybrid genotypes in the progeny should have marker genotypes consistent with those of the two parent plants (RUIZ et al., 2000).

To better investigate the organization of the citrus genome and locate markers for valuable traits, many citrus genetic maps have been developed during the past years, most of which

were composed of random amplified polymorphic DNA (RAPD), amplified fragment length polymorphism (AFLP) and inter-simple sequence repeat (ISSR) markers, and the rest by restriction fragment length polymorphism (RFLP), simple sequence repeats (SSR), sequence characterized amplified region (SCAR), cleaved amplified polymorphic sequence (CAPS), sequence tagged site (STS) and single nucleotide polymorphism (SNP) markers (CHEN et al., 2008). The recent five citrus genetic maps (male and female Nules Clementine, Chandler pummelo, Pink pummelo and sweet orange) were established with codominant markers (primarily SSR markers) and sequences in public available databases, based on haploid Clementine genome sequence produced by International Citrus Genome Consortium (ICGC) (OLLITRAULT et al., 2012). The citrus genetic maps are utilized for map-based cloning (MBC) and marker-assisted selection (MAS) (CHEN et al., 2008). Since more informative markers are required to build a consensus reference map, integration and comparison of genetic and physical maps are essential for genome sequencing and comparison of the citrus genome in different species and hybrids (CHEN et al., 2008). Ruiz and Asins firstly reported the comparison of five linkage maps of *Citrus* and *Poncirus* parents based on linked SSRs, IRAPs, RAPDs, RFLPs, isozymes and other markers (RUIZ AND ASINS, 2003). Molecular markers have been widely used in citrus cultivar identification and mapping. Sequence-related amplified polymorphism (SRAP), SSR, ISSR, peroxidase gene polymorphism (POGP), resistant gene analog (RGA), RAPD and *Cabsr* gene were used to construct the ‘Clementine’ mandarin (*Citrus reticulata* Blanco ‘Clementine’) and ‘Orlando’ tangelo (*C. paradisi* Macf. ‘Duncan’ × *C. reticulata* Blanco ‘Dancy’) genetic linkage maps (GULSEN et al., 2010). A saturated genetic map of

Clementine including 961 markers was developed by the International Citrus Genome Consortium (ICGC) (OLLITRAULT et al., 2012). This map greatly facilitated the whole genome sequence assembly of a haploid from Clementine (OLLITRAULT et al., 2015; WU et al., 2014). Due to the advantage of SNP markers for high throughput methods, SNP markers can be used efficiently for comparative mapping, and thus a GoldenGate® SNP array was developed (OLLITRAULT et al., 2015). A highly conserved linear order of markers between pummelo and sweet orange was shown by using comparative mapping with Clementine reference genome (OLLITRAULT et al., 2015). High-density mandarin genetic maps developed using an Illumina GoldenGate® SNP array were constructed by Gmitter et al., which associated fruit quality traits with SNP markers for MAS in mandarin breeding (YU et al., 2016). The high-density pummelo genetic map including 1543 SNP and 20 SSR markers was constructed by using restriction site-associated DNA (RAD) sequencing and designed SNP markers, providing a prerequisite for MAS breeding and pummelo genome assembly (GUO et al., 2015). For citrus germplasm diversity analysis, SSR markers were utilized to detect molecular polymorphisms and phylogenetic relationships between *Citrus* accessions and *Citrus* relatives, due to their highly polymorphic and codominant features (BARKLEY et al., 2006). In other studies, polymorphisms of molecular markers were extensively exploited to measure phylogenetic relationships of citrus and related genera (AMAR et al., 2011; FEDERICI et al., 1998).

In plants, SNP markers and Indel markers are most commonly used for genomic selection, genomic association and QTL mapping due to their abundance and co-dominance in the genome (YAMAKI et al., 2013). Compared to traditional marker systems, the SNP and Indel

markers can be easily generated within even a small region of the genome, and the genotyping procedure is fast, dependable and economical (HAYASHI et al., 2004). Compared to SSR markers, SNP markers are more stable with low mutation rate and amenable to automation and high-throughput analysis such as large-scale genotyping systems (JIANG et al., 2010; YAMAKI et al., 2013). The simplest but low-throughput method is by an electrophoresis-based assay, in which SNP markers can be genotyped by allele-specific PCR and Indel markers can be genotyped by PCR amplicon size polymorphisms (HAYASHI et al., 2006). In studies on nine rice blast resistance genes, Hayashi et al. utilized segregation information from progeny in crosses and published gene resources to develop a marker system with SNP markers that flanked and were closely linked to the target genes (HAYASHI et al., 2006).

Being the most frequently occurring genetic variation in the genome, SNP can be used not only for the relating sequence variations to phenotypic changes, but also genotyping studies and further phylogenetic studies. Although genotyping using SNPs with multiplex assays has become more practical, simple and rapid methods are also available when genotyping a single or a few SNPs of interest. In genomic DNA, both SNP alleles must have at least 1% allele frequency to be considered as a SNP at the population level (VIGNAL et al., 2002). Although in theory, four possible nucleotide bases can be present at each position, in practice SNPs are usually bi-allelic, due to both the low frequency of single nucleotide substitutions at the origin of SNPs and the bias in mutation (VIGNAL et al., 2002). Mutation can cause either transition exchanges (purine-purine or pyrimidine-pyrimidine) or transversion exchanges (purine-pyrimidine or pyrimidine-purine), which is theoretically

expected to occur twice as often as transitions under random mutation situations but a clear bias towards transitions is observed in actual data (VIGNAL et al., 2002). Traditionally, SNP genotyping can be performed with multiple techniques, such as direct hybridization, restriction enzyme cutting, primer extension, oligonucleotide ligation assay, exonuclease detection, invasive cleavage of oligonucleotide probes and single strand DNA conformation and heteroduplexes, and others (VIGNAL et al., 2002). Recent SNP genotyping is performed in a high-throughput way, such as SNP array, reduced-representation sequencing, RAD-seq, low-coverage sequencing and whole genome sequencing (DAVEY et al., 2011). In addition, SNPs can be easily converted into other conventional molecular markers, including RFLP, CAPS and dCAPS, providing further convenience to situations where SNP genotyping is unavailable (JIANG et al., 2010). In citrus, a large SNP database is available from citrus sequencing projects, and with Illumina GoldenGate assay, over 1,000 SNP markers were mapped onto the Clementine reference genome (CUENCA et al., 2013; YU et al., 2016).

Conventionally, a specific target SNP can be genotyped according to the presence or absence of the PCR products on an agarose gel by using allele-specific PCR primers, by which method the allele-specific PCR could be useful for map-based cloning and marker-assisted selection in crops (HAYASHI et al., 2004; HAYASHI et al., 2006). Besides low-throughput, the drawback of allele-specific PCR for SNP markers is that a single nucleotide mismatch at the 3'-end of the nonspecific allele is may not be sufficiently reliable to discriminate between two alleles (HAYASHI et al., 2004). By using a modified allele-specific PCR method, Hayashi et al. confirmed 67% of the primers initially generated had

sufficient specificity in discriminating alleles, indicating the reliability of this method (HAYASHI et al., 2004). However, in some situations, a single nucleotide difference at the 3'-end of the nonspecific allele is not sufficient to reliably discriminate two alleles (DRENKARD et al., 2000).

Small insertion/deletions (Indels) are one of the major alternative forms of genetic variation that frequently occur across the genomes of diverse species and account for a significant part of sequence divergence between closely related species in animals, plants, insects and bacteria (DAS et al., 2015). Here we use the term "Indel" to refer to small insertion/deletions in sequences that are not repetitive, distinguishing them from SSR. Indels originate from the insertion and excision of retrotransposons or mobile elements, unequal crossover events or slippage in simple sequence replication (BRITTEN et al., 2003). Therefore Indel mutations seldom occur at exactly the same genomic position with the same length (GARCIA-LOR et al., 2012). Similar to SNPs, Indels are also highly abundant and distributed throughout the genome, and can serve as molecular markers in marker-assisted selection and genetic mapping (HAYASHI et al., 2006). In *Drosophila melanogaster* and *Caenorhabditis elegans*, Indels represent between 16% to 25% of all genetic polymorphisms, and the proportion in human is similar (MILLS et al., 2006). Also, Indels can be responsible for functional changes if they are in coding regions, so they are considered as an essential source for evolutionary change (GARCIA-LOR et al., 2012). In human, Indels are the major source of gene defects (BRITTEN et al., 2003). Indels were successfully used for genetic studies of wheat, rice and natural populations (HAYASHI et al., 2006; RAMAN et al., 2006; VÄLI et al., 2008). Indel markers were also proved to be a

highly reliable, informative and cost-effective alternative marker for variety testing, diversity analysis and mapping, in Steele's studies on distinguishing Basmatis from other fragrant rice varieties, as well as splitting Basmatis and other varieties into fragrant and non-fragrant varieties groups (STEELE et al., 2008). In the Indel analysis for sweet orange and satsuma, the average Indel frequency was one every 1030bp, the majority of which were four bases or less in length with the number of Indels decreasing exponentially according to their length (JIANG et al., 2010). In the first specific molecular and phenotypic studies of citrus varieties using Indels as molecular markers, 12 Indel polymorphisms were identified in 10 genes including primary and secondary metabolite biosynthetic pathways that determine citrus fruit quality and salt tolerance, in which some are specific to *C. medica* and *C. maxima* (GARCIA-LOR et al., 2012). Compared to SSR markers, although Indel markers are less polymorphic and less informative at the intraspecific level, Indel markers have higher interspecies differentiation and are better in tracing the contributions of three ancestral citrus species to the genomes of modern varieties than SSR markers in phylogenetic studies (GARCIA-LOR et al., 2012; OLLITRAULT et al., 2014). Also, frequent homoplasy demonstrated in citrus limits the value of SSR markers for phylogenetic analysis (BARKLEY et al., 2009; OLLITRAULT et al., 2014). The Indel markers developed from sequence data not only provide a valuable molecular resource for germplasm diversity analysis, but also contribute to deciphering the interspecific mosaic structure of secondary cultivated citrus and analyzing intraspecific diversity (OLLITRAULT et al., 2014). The discovery of Indel markers makes the selection of specialized markers with high phylogenetic diagnosis potential analysis more efficient (OLLITRAULT et al., 2014).

Therefore, Indel markers are suitable for intra-population diversity analysis and useful for small-scale studies in which they can be easily analyzed in small laboratory structures (GARCIA-LOR et al., 2012). The only limitation is the relatively low frequency and some pluri-allelism of Indel markers other than SSR markers (OLLITRAULT et al., 2014). As a result, to bridge the gap between SNPs and other polyallelic molecular markers, the identification and development of more polymorphic Indel markers become attractive.

Compared to SNPs, Indels are responsible for more unmatched nucleotides in genomes than single base substitutions, and thus can be discovered by PCR product length differences following amplification of a small genome region that contains the insertion/deletion (BHATTRAMAKKI et al., 2002; DRENKARD et al., 2000; GARCIA-LOR et al., 2012). Technically, genotyping using Indels can be performed by size separation after targeted PCR (VASEMAGI et al., 2010). With the assistance of a significant amount of sequence information obtained by next-generation sequencing, PCR-based Indel markers can be developed and become an essential and inexpensive source of molecular markers for citrus phylogenetic and genetic association studies (GARCIA-LOR et al., 2012). One of the limitations on using Indel markers is that their frequency in genomes is lower than that of SNPs, and the nature of binary variation of Indel marker carries lower informative power and poor biostatistical power per locus compared with markers such as SSRs; therefore this lower power has to be compensated by using a more substantial number of markers per case (DRENKARD et al., 2000; HAYASHI et al., 2004; HUANG et al., 2014). By combining multiple Indel markers, not only is the discrimination power significantly improved, but also the time and cost can be reduced dramatically (HUANG et al., 2014). The advantage of

Indels over SNPs is that they can be examined by directly measuring the length difference between alleles due to their length polymorphisms especially when the amplicon sizes are relatively small. When fluorescent-labelled primers and a capillary sequencer are available, PCR products differing in length by as little as one or two base pairs can be distinguished (DAVIS AND HAMMARLUND, 2006). Differences in PCR product length of insertion/deletion are usually minimal and thus the standard agarose gel electrophoresis may not be sufficient for discrimination of PCR products corresponding to particular Indels (DRENKARD et al., 2000). In Yamaki's studies, 22 PCR-based Indel markers were developed to discriminate 22 genotypes of genus *Oryza*, proving the usefulness and reliability of Indel markers in identifying germplasm (YAMAKI et al., 2013). However, PCR and agarose gel electrophoresis are not a high throughput method, and therefore these methods of Indel analysis are less useful for analyses dependent on a large number of markers.

The feasibility of using real-time quantitative PCR (qPCR) technology for breeding and chimerism monitoring is demonstrated by several studies, and it can provide a fast, sensitive, and elegant quantification of genotypes (GARRITANO et al., 2009). The availability of qPCR molecular marker sets for every parent-mutant pair is a must for routine chimerism monitoring (GARRITANO et al., 2009). qPCR has been widely used in research and diagnostic applications. For plant genomic studies, qPCR is a valuable tool to determinate gene copy numbers, identify transgenic plants, and to genotype SNPs, Indels, and other structural rearrangements (SUN et al., 2010). Among various techniques, high-resolution melting curve (HRM) is a recent advance in the detection of SNPs and Indels (GARRITANO et al., 2009; SHIOKAI et al., 2009). However, the primary challenge is that for

many species the number of SNPs and Indels published to-date for PCR or qPCR is too limited to guarantee the applicability for every parent-mutant pair. Also, HRM is not a high throughput method.

The availability of high-throughput sequencing data empowers the analysis of multiple sequences alignment with considerable redundancy, which facilitates the discovery of potential SNPs and Indels with high confidence and authenticity and eliminates fallacious SNPs or Indels that are caused by random or systematic sequencing errors (JIANG et al., 2010). Many high-throughput methods, such as expressed sequence tags (ESTs), have been developed to achieve the requirements on SNP discovery of citrus varieties (FUJII et al., 2013). To interpret the structure and the phylogeny of citrus genomes, a reference genome sequence of a haploid clementine was generated through the International Citrus Genome Consortium (ICGC) (WU et al., 2014).

Citrus taxonomy used to be controversial because so many taxa are inter-fertile. The development of molecular markers has resulted in a new view of how citrus types arise and maintain their integrity, as well as the relationships between various *Citrus* types, which showed that most cultivated types arose from three ancestral species (MOORE, 2001). In order to better discriminate citrus varieties, a significant amount of resources is currently available to develop markers that are suitable for distinguishing closely related cultivars, including large numbers of expressed sequence tags (ESTs) from the National Center for Biotechnology Information (NCBI) and SNP arrays (JIANG et al., 2010). The development of molecular markers within and among citrus accessions is also essential in citrus genome-wide association studies, breeding and germplasm diversity analysis (JIANG et al., 2010).

SNPs were used to distinguish cultivars from two closely related citrus and developed into a marker system for citrus that showed the capacity to distinguish citrus genotypes and evaluate the level of polymorphism (NOVELLI et al., 2004). The utilization of the available sequences in public databases to discover molecular markers was proved to be economical and dependable in many studies in different plant and animal species (JIANG et al., 2010). SNP arrays have been developed and extensively applied for genotyping *Citrus* cultivars and high-density genetic map creation. The Affymetrix citrus GeneChip® was developed based on EST sequencing projects, which contains 30,264 unigene probe sets and supported whole genome sequence alignment (CLOSE et al., 2006; GMITTER JR et al., 2007; XU et al., 2013). For future genome-wide association studies in citrus, the development of high-density molecular marker systems for citrus becomes essential.

W. Murcott mandarin, also known as Nadorcott or Afourer, is originally from Morocco and was imported to California in 1985 (WU et al., 2014). W. Murcott is considered to be a complex hybrid of a chance zygotic seedling from an uncertain pollen parent and Murcott tangor, which is presumed an F1 hybrid of a cross between an unknown mandarin and a sweet orange (WU et al., 2014). W. Murcott is a mid-late season mandarin whose fruits are deep orange, sweet juicy and easily peelable, and can be seedless (ROOSE AND WILLIAMS, 2007). Because of their high fruit quality, W. Murcott trees are widely planted throughout California. The seedlessness of W. Murcott is due to its self-incompatibility, in which the male gametes are unable to fertilize female gametes, preventing seed formation and resulting in facultative parthenocarpy (GAMBETTA et al., 2013; OLLITRAULT et al., 2007). However, in order to set seedless fruits, W. Murcott trees are required to be either planted

in isolated areas without other pollen-bearing citrus or covered with screen during flowering season, otherwise seedy fruits can be produced as a result of cross-pollination by bees-transferring pollen from other Citrus with high pollen viability, which could significantly reduce the fruit value (CROWLEY, 2011).

Tango mandarin is a low-seeded selection of W. Murcott (Afourer) mandarin, which may be the same as Nadorcott. Budwood of W. Murcott from Lindcove, California was treated with 50 Gray units of ^{60}Co gamma at Riverside, California and the irradiated buds were propagated on various rootstocks in a greenhouse and then planted in the field for fruit production and evaluation (ROOSE AND WILLIAMS, 2007). One tree propagated on 'C32' citrange rootstock set fruits with very low seed counts in all situations of cross-pollination and very similar fruit quality and production as on the original W. Murcott cultivar (ROOSE AND WILLIAMS, 2007). This tree was then selected for further trials in which buds were propagated onto Carrizo and C35 citrange rootstock (ROOSE AND WILLIAMS, 2007). Multiple rounds of propagation were performed before this cultivar was named Tango and released for commercial production.

It was considered that most radiation-induced mutations are DNA deletions, as a result of repair or misrepair of induced DNA double-strand breaks (DSBs) (SANKARANARAYANAN et al., 2013). For example, gamma irradiation caused a large deletion mutation with breakpoint possibly near short direct or inverted repeats, and insertions of hundreds of bp of repetitive sequences in hamster gene (MILES et al., 1990). The error-prone DSB repair and DNA replication-based mechanisms are responsible for deletions and structural variations in the genome (SANKARANARAYANAN et al., 2013). The origin of tandem repeat

polymorphisms, Indels and mobile element insertions is through two families of DSB repair pathways, non-homologous end-joining (NHEJ) and homologous recombination repair, and a replication-associated process, microhomology-mediated break-induced replication (MMBIR) (SANKARANARAYANAN et al., 2013). NHEJ, microhomology-mediated end-joining (MMEJ), non-allelic homologous recombination (NAHR) are error-prone. Most radiation-induced DSBs cannot be directly ligated due to their complexity. In order to make the two DNA ends into a ligatable configuration, some processing and gap-filling are required, during which insertions or deletions of a few nucleotides are introduced at the junctions in NHEJ repair (SANKARANARAYANAN et al., 2013). During MMEJ repair, 5-25 bp microhomology is required for alignment of broken ends before joining and this minimal homology can result in deletions flanking the breakpoint (SANKARANARAYANAN et al., 2013). NAHR repairs the homologous chromosomes between misaligned repetitive sequences, during which process the sequences between repeats are either deleted or duplicated, resulting in the copy number change (SANKARANARAYANAN et al., 2013). Almost all naturally occurring deletions of different sizes can be explained with NHEJ, MMEJ, NAHR, MMBIR and strand slippage during replication (SANKARANARAYANAN et al., 2013).

Although DNA repair processes in plants are less well characterized than in humans, similar mechanisms occur and since Tango is derived from gamma irradiation of budwood of W. Murcott, it is likely that deletions and chromosome structural changes are involved and can be used as markers to differentiate the mutant from the parent. The objective of this study is to develop fast, reliable, and low-cost PCR-based molecular markers and

validate a new allele-specific system for discrimination of W. Murcott mandarin and Tango mandarin to supplement the already existing phenotype markers and improve the applicability of PCR or qPCR for a more sensitive, accurate and economic method to distinguish between the two varieties. We designed primer sets specific for eight biallelic SNP/Indel loci and validated them in several W. Murcott mandarin trees and Tango mandarin trees. The results demonstrate the feasibility of a SNP/Indel-based qPCR approach for rapidly distinguishing Tango from W. Murcott, and provide a path toward an easy-to-implement method for better Tango patent protection.

Materials and Methods

Tissue Collection and DNA Extraction

All tissue samples were young leaf tissue of W. Murcott trees and Tango trees growing at the Citrus Research Center, University of California, Riverside. Leaf samples were collected from 5-8 different trees of Tango and W. Murcott. DNA samples were extracted from leaves as follows. Leaves were cleaned with cold tap water and air dried. After washing, leaves were ground into fine powder in liquid nitrogen with a mortar and a pestle. The pulverized tissue was kept frozen on dry ice during grinding of other tissue samples and later transferred and stored in a -80 ℃ freezer until extraction. For DNA extraction, approximate 100 mg frozen tissue was extracted using Qiagen DNeasy Plant Mini Kit following the manufacturer's instructions except that at step 19: instead of an additional 100 µl of Buffer AE being pipetted onto the DNeasy membrane, ~100 µl of the solution that had just passed through the membrane was added, and centrifuged into a microcentrifuge collecting tube, in order to enhance the concentration of extracted DNA. After DNA extraction, the DNA concentration was measured with a Nanodrop 2000 spectrophotometer. This DNA concentration in ng/µl was used to calculate the amount of DNA used in qPCR reactions.

Sequencing of W. Murcott Mandarin and Tango Mandarin

The sequencing data used in this study was Illumina paired-end libraries prepared from genomic DNA of a W. Murcott mandarin tree in the UCR Citrus Variety Collection and a tree in field 1B propagated directly from the original Tango "mother tree" at the University of California, Riverside. Genomic DNA was mechanically fragmented by sonication to fragment lengths of approximately 300 bp followed by size selection. The size selected DNA fragments were purified from E-Gel® Agarose Gel Electrophoresis System and checked for fragment length distribution on Agilent DNA 7500 chips (Agilent Technologies, Germany). Paired-end libraries for W. Murcott and Tango respectively were prepared according to the manufacturer's instruction (Illumina "Preparing Samples for Paired-End Sequencing", Part # 1005063 Rev. A June 2008). One lane of each library was sequenced on an Illumina Genome Analyzer II at the Institute for Integrative Genome Biology at the University of California, Riverside. The average insert size for both libraries is approximately 275 bp.

Also, Tango and Nadorcott were sequenced in IVIA Spain. Analysis of the sequences from IVIA Spain identified eight mutations in Tango, which includes two short deletions, DC-1 and DC-7, and six SNPs, 1-GP, 3-GP, 5-GP, 7-GP, 33-GP and 45-GP. Among the eight mutations, three SNPs (3-GP, 5-GP and 33-GP) and two short Indels (DC-1 and DC-7) were observed in the sequence data generated at UCR, the other three SNPs (1-GP, 7-GP and 45-GP) were not observed. We thank Manuel Talon from IVIA for providing mutation information on the sequences.

Sequence Analysis

W. Murcott and Tango raw data in FASTQ format generated from Illumina Genome Analyzer II was trimmed, filtered and aligned against the haploid *Citrus clementina* 1.0 genome assembly (composed of nine long scaffolds, corresponding to nine chromosomes, and one short scaffold which could not be assembled with any of the nine pseudochromosomes) with Burrows-Wheeler Aligner (BWA) by Prognosys Biosciences. The minimum mapping quality of reads was set to 35, resulting in approximately 119 million aligned reads for W. Murcott and 117 million aligned reads for Tango, and an average depth of coverage of approximately 40X for both W. Murcott and Tango data.

SNP and short Indels were analyzed based on aligned reads data in bam format of W. Murcott and Tango, and *Citrus clementina* scaffolds as the reference genome. The duplicate reads of all alignment data were firstly marked and removed by Picard v2.14.1 (http://broadinstitute.github.io/picard) with Picard MarkDuplicates, where parameters were set as follows:

```
java -jar –Xmx8g picard.jar MarkDuplicates INPUT=input.bam
OUTPUT=output.bam METRICS_FILE=metrics
REMOVE_DUPLICATES=true
```

Two programs, Pindel v1.0 (http://gmt.genome.wustl.edu/packages/pindel) and Dindel v1.01 (https://sites.google.com/site/keesalbers/soft/dindel), were used to find potential Indel loci respectively. For Pindel, the default settings were used:

```
./pindel -f Clem1.0.fa -i config.txt -c ALL -o output
```

where the average insert size of both W. Murcott data and Tango data was set to 275 bp in the configuration text file.

For Dindel, the default settings were used:

```
./dindel-1.01-linux-64bit --analysis getCIGARindels --
bamFile input.bam --outputFile dindel_output --ref
Clem1.0.fa

python selectCandidates.py -i dindel_output.variants.txt -
o output.variants.txt --minCount=5

python makeWindows.py --inputVarFile output.variants.txt -
-windowFilePrefix output.realign_windows --
numWindowsPerFile 10000

./dindel-1.01-linux-64bit --analysis indels --doDiploid --
bamFile input.bam --ref Clem1.0.fa --varFile
realign_windows.txt --libFile dindel_output.libraries.txt
--outputFile dindel_output_windows

python mergeOutputDiploid.py --inputFiles
dindel_outputfiles.txt --outputFile variantCalls.VCF --ref
Clem1.0.fa
```

Then the loci that appeared in both results were selected for additional filtering. Within the 16,525 chosen short Indels, multiple filtrations were performed to reduce the pool of potential short Indels. First, all the results that did not pass the quality filter were removed. Second, the Indel position should be heterozygous in Tango and homozygous in W. Murcott. Tango should have one haplotype with the Indel and one without it, while for W. Murcott, either both haplotypes contain the Indel or no haplotype contains the Indel, without supporting reads for the other haplotype. Therefore, the predicted Indel positions where: 1. W. Murcott is heterozygous and Tango is homozygous, and 2. W. Murcott is mostly homozygous with at least one Indel haplotype read when Tango is heterozygous, were removed. That means in the output VCF file, the loci with GT (genotype) = 0/1 in Tango, GT = 0/0 in W. Murcott and AD (allele depth, how many reads support this allele) = 0 in W. Murcott were kept. After filtration, 117 predicted potential Indel loci were retained. The 117 predicted Indels were then manually examined for read alignment using Integrative Genomics Viewer (IGV, ver. 2.4) (http://software.broadinstitute.org/software/igv/) and filtered to remove predicted Indels for which the region in they reside contained more than one Indel haplotype in W. Murcott or more than two Indel haplotypes in Tango. Finally, Indel loci around which the sequence was suitable to design real-time PCR primers and conventional PCR primers were retained. In addition to two short Indels that were reported by Dr. Manuel Talon from IVIA, 38 more short Indel loci were selected to be validated using real-time PCR. To prioritize the Indels to be tested, the Dindel genotype quality of each predicted Indel was used, with values greater than or equal than 100 selected. This subset, which included 19 Indels with

genotype quality values from 100 to 889, was evaluated by real-time PCR. Other predicted Indels with genotype quality values lower than 100 were not tested.

Allele-Specific PCR Primer Design

The molecular markers were developed based on the Tango and W. Murcott DNA sequences. Allele-specific primer sets were designed based on the Tango-specific sequence to target Indels or SNPs that were only present in heterozygous form in Tango and for which one allele was absent in W. Murcott. For SNPs, the primers were manually designed by placing the nucleotide unique to a Tango haplotype at the 3' end of one primer, combined with a second primer whose sequence was present in both alleles in W. Murcott and Tango (GAUDET et al., 2009). For Indels, similarly, the primer sequences were designed from a sequence inferred by concatenation of the sequences flanking the deletion allele, combined with a second primer whose sequence was present in both alleles in W. Murcott and Tango. By the above strategy, the allele-specific primer sets only amplify the Tango-specific allele. Also, SSR markers (OLLITRAULT et al., 2012) were used as background control. Primer design and sequences are summarized in Figure 2.1 – 2.5 and Table 2.1.

Allele-Specific PCR Amplification

Reactions included: 1 mM GoTaq buffer, 0.2 mM dNTPs, 2.0 mM $MgCl_2$, 0.05 U/µl GoTaq® DNA polymerase (Promega), 2 ng/µl W. Murcott or Tango genomic DNA, 0.2 µM forward primer, 0.2 µM reverse primer, and ddH_2O to 10 µl total reaction volume. The reactions were placed in a heat-enabled lid thermocycler programmed with the following settings: For DC-1 deletion: initial 95 ℃ step for 30 sec; 95 ℃ denaturation step 10 sec; 54 ℃ annealing step 30 sec; 72 ℃ elongation step 30 sec; repeat denaturation, annealing, and elongation steps 34 times; 72 ℃ 5 min; final 4 ℃ indefinite cooling step. For DC-7 deletion: initial 95 ℃ step for 30 sec; 95 ℃ denaturation step 10 sec; 68 ℃ annealing step 30 sec; 72 ℃ elongation step 30 sec; repeat denaturation, annealing, and elongation steps 34 times; 72 ℃ 5 min; final 4 ℃ indefinite cooling step. For 5-GP SNP: initial 95 ℃ step for 30 sec; 95 ℃ denaturation step 10 sec; 65 ℃ annealing step 30 sec; 72 ℃ elongation step 30 sec; repeat denaturation, annealing, and elongation steps 34 times; 72 ℃ 5 min; final 4 ℃ indefinite cooling step. During thermocycling, an agarose gel containing 1% agarose and 0.03% (v/v) 10,000X diluted GelRed™ nucleic acid gel stain (Biotium Cat: 41003) was prepared. After PCR amplification completion, PCR products were analyzed on the agarose gel to determine that if PCR fragments of the expected size were present. 1 µl 6X loading dye was mixed with 10 µl of every sample. The gel was run at 100V for 1 hour. Amplicon sizes were estimated by comparison with a 100-bp DNA ladder.

HRM Real-Time PCR Primer Design

Since some Indels were relatively short (1-3 bp), HRM real-time PCR was utilized instead of conventional PCR to obtain a better resolution than offered by an agarose gel. Due to the sensitivity of HRM real-time PCR, primer design was critical for success. For each Indel locus, up to five real-time PCR primer pairs were designed using IDT PrimerQuest Tool (https://www.idtdna.com/primerquest/Home/Index). For Customer Design Parameter, "Primer Tm (°C)" was set to between 62 °C minimum to 65 °C maximum with optimized Tm at 62 °C, and "Amplicon Size (bp)" was set between 75 bp minimum to 150 bp maximum without optimized size. The Indel position relative to the amplicon was set at "Included Region" field, making the designed primers flank the Indel-containing DNA fragment. 52 pairs of primer were designed in total and 7 best pairs were selected, including DC-1 and DC-7. Figures 2.6 – 2.12 and Table 2.2 shows the primer sets used for analysis of seven short Indels on different scaffolds.

HRM Real-Time PCR Amplification

HRM real-time PCR was performed and analyzed with ABI QuantStudio. The reaction mixture contained diluted DNA, 1X MeltDoctor HRM Master mix and 10 uM of each primer in a total volume of 20 uL. The real-time PCR cycle profile was: 10 min of an activation period at 95 ℃, followed by 40 cycles of a two-step profile involving 15 sec at 95 ℃ for denaturation and 1 min at 60 ℃ for annealing and extension. The melt analysis conditions were 10 sec at 95 ℃, followed by 1 min at 60 ℃, and a ramp from 60 ℃ to 95

℃, rising by 0.25 ℃ every 5 sec, and a final 15 sec at 60 ℃. Melt curves and derivative melt curves were plotted for each pair of primers using Applied Biosystems QuantStudio™ 6 & 7 Flex Real-Time PCR System. The extra real-time PCR product present in Tango mandarin relative to W. Murcott mandarin was identified using the same software. To test the stability of the real-time PCR marker, five leaf samples from five different W. Murcott mandarin trees and five leaf samples from five different Tango mandarin trees were analyzed.

TA Cloning and Sanger Sequencing

Seven regions of the citrus genome were amplified by PCR and used for cloning and sequencing. The regions chosen are known to have the predicted Indels in different haplotypes in Tango. This information is based on aligned reads of Illumina NGS data generated in this study. Primers were designed using IDT PrimerQuest (https://www.idtdna.com/PrimerQuest/) to allow amplifying fragments of about 500 bp - 750 bp in length by PCR using Taq DNA Polymerase (New England BioLabs®) (Table 2.3). The amplified products will have a non-templated A added to the 3' terminus. TA cloning was then performed using a Topo® TA Cloning® kit (Invitrogen) according to the protocol suggested by the manufacturer. Briefly, the PCR products were electrophoresed on a 0.8 % agarose gel, DNA bands were excised, purified using QIAEX II Gel Extraction Kit (Qiagen), pure DNA fragments were ligated to the Topo® vector at room temperature for 5 minutes. The ligated product was incubated with chemically competent *Escherichia*

coli cells supplied with the cloning kit, subjected to heat shock treatment for 45 seconds at 37 ℃. About 500 ul of Luria-Bertani (LB) broth was added to the mix and the transformed cells were grown for 45 minutes at 37 ℃ with shaking at 300rpm. The bacteria were then plated on petri dishes containing solid LB medium with 100 ug/ml ampicillin for selection, grown overnight at 37 ℃ and the colonies were stored in the refrigerator.

The rolling circle amplification (RCA) method was used for sequencing. The bacterial colonies were grown in 96-well plates containing liquid LB and 100 ug/ml ampicillin overnight at 37 ℃. Plates were covered with sealing film and shipped to Molecular Cloning Laboratories (MCLAB) at San Francisco, CA for sequencing. In RCA method, the circular plasmid DNA is amplified to produce sequencing templates. Amplified templates are then sequenced using M13 forward and M13 reverse primers. Sequences were trimmed to removed low-quality bases and then aligned to the *Citrus clementina* 1.0 reference genome.

Results

Allele-Specific PCR Amplified Tango-Specific Products for Five of Eight Candidate Markers

BLAST search of sequences of primers including the mismatch nucleotide (the allele-specific primer in a primer pair for SNP) or concatenated sequence over the deletion (the allele-specific primer in a primer pair for Indel) against *Citrus clementina* 1.0 reference genome on Phytozome (https://phytozome.jgi.doe.gov) did not show any full-length match with 100% identity, while sequences of the other primer in a primer pair (non-allele-specific primer) showed a unique hit at the target site, or multiple hits including target site and sites that were far from target sites or on a different chromosome, indicating the full-length sequence of the putative mutant allele does not exist in the reference genome, and the PCR products, if any, are less likely to be amplified from template sequences outside of the target region of interest.

Allele-specific PCR results showed both W. Murcott and Tango samples had the expected SSR marker product, indicating that the PCR reaction worked for both W. Murcott and Tango DNA. The products that were only present in the Tango samples had the size expected from the Tango genome sequence and the allele-specific primer design set. The extra PCR product in one of the W. Murcott samples with the same size as in Tango PCR allele-specific PCR products may be due to contamination with Tango DNA. DC-1 deletion showed both W. Murcott leaf DNA samples and Tango leaf DNA samples had a product

of approximately 130 bp, which corresponded to the control SSR marker CF-ATG05 product. Also, Tango leaf samples had an additional product with size approximately equal to the predicted 303 bp allele-specific primer product. Similarly, PCR results for DC-7 short deletions, 3-GP, 5-GP and 33-GP SNPs showed both W. Murcott leaf DNA samples and Tango leaf DNA samples had products of the approximate size expected for either CF-ATG05 or CF-AT01 control SSR marker, and Tango leaf samples had an additional product approximately the size expected for the allele-specific primer set (Figure 2.13 - 2.17). The other three SNPs, 1-GP, 7-GP and 45-GP, produced the same products with W. Murcott and Tango template DNA and the allele-specific primer sets (data not shown).

Most samples produced a product of the expected size, but some of them failed to amplify or had an unexpected product which may be due to contamination. For DC-1, one Tango sample failed to amplify with both SSR primers and allele-specific primers, and another Tango sample failed to amplify with allele-specific primers. One of the W. Murcott samples showed an extra product in DC-1, DC-7, 3-GP and 5-GP with a size similar to the product amplified by the allele-specific primer set in Tango, a result which may be due to contamination with Tango DNA. One W. Murcott sample failed to amplify with DC-1, another W. Murcott sample failed to amplify with 5-GP and a third W. Murcott sample failed to amplify with 33-GP (Figure 2.13 - 2.17).

HRM Real-Time PCR Showed Different Derivative Melt Curves Between W. Murcott and Tango

BLAST search of each sequence of HRM real-time primer pairs flanking the short deletion against *Citrus clementina* 1.0 reference genome on Phytozome (https://phytozome.jgi.doe.gov) also showed an unique hit at the target site, or multiple hits including target site and at sites that were far from target sites or on a different chromosome, indicating the PCR products, if any, are likely to be from the target region of interest.

Since real-time PCR is very sensitive and can detect differences as small as a single base between sequences, we can predict that the small Indel differences between W. Murcott and Tango can be identified from the melt curve. The different amplicons will have different melt peaks. The aligned melt curves and derivative melt curves of all the loci clearly showed differences between the curves of W. Murcott and Tango, except Del3-3 did not show a convincing difference (Figure 2.18 - 2.24). The differences were due to the slightly different melting temperatures of double-stranded DNA of different amplicons of W. Murcott and Tango from the same primer pair, in which W. Murcott had only one product without small deletions while Tango had a mixed product including sequences with and without small deletions. In the aligned melt curve, the difference was evident as the "bubble" between W. Murcott and Tango melt curves during the fast fluorescence strength decreasing phase. In the derivative melt curve, the difference was embodied as the melt peak present at a different temperature or as a different number of melt peaks between W. Murcott and Tango. For W. Murcott DNA samples in real-time PCR, no sample showed

characteristics suggestive of contamination with Tango. All five Tango samples (red curves) were different from all five W. Murcott samples (green curves), while no amplification for NTC (blue curves), indicating Tango had different amplicons than W. Murcott for the same pair of primers, which satisfied the requirement for effective molecular markers.

Sanger Sequences of TA Clones Confirm Heterozygous Deletions and SNPs

PCR products of DC-1 using primers flanking the target short deletion were cloned and 18 clones of Tango were sequenced. By aligning the sequences against the Clementine reference genome, two haplotypes were observed within the 18 clones from Tango: one haplotype with 10 clones missing the 20-nt sequence GCTACTTATAGACTCTTAGT, while the other haplotype with eight clones containing this sequence. The BLAST result of the 20-nt showed it is located on scaffold 1 from 19024294 bp to 19024313 bp according to *Citrus clementina* genome assembly 1.0, while W. Murcott carries no such deletion on either chromosome (data not shown). Also, some nucleotides near the short deletion were observed to be different between the two haplotypes, and the SNPs are linked with the deletion. The allele with missing nucleotides linked with some SNPs upstream and downstream. Upstream of the allele with the deletion there is a T at 19024287 bp and C at 19024290 bp, while those two nucleotides were replaced by G and T respectively on the other allele. Downstream of the deletion, the closest three SNPs are T at 19024315 bp, G at 19024326 and T at 19024334, which were replaced by C, A and C on the allele without

deletion respectively. This haplotype structure is consistent with the deletion being allelic with the intact allele.

DC-7 was confirmed as a 13 bp heterozygous deletion in Tango located on scaffold 7 from 6988443 bp to 6988455 bp of *Citrus clementina* genome assembly 1.0 with sequence TTGGTGAAGAAAA, while W. Murcott carries no such deletion on either chromosome. 3-GP is a transition from C to T in one of the two alleles of Tango at 6204314 bp on scaffold 1, while both alleles of W. Murcott have C at the same position. 5-GP is a transversion from G to C in one of the two alleles of Tango at 19505366 bp on scaffold 1, while both alleles of W. Murcott have G at the same position. 33-GP is a transversion from A to T in one of the two alleles of Tango at 2717333 bp on scaffold 7, while both alleles of W. Murcott have A at this position.

Other newly discovered short deletions are all heterozygous deletions in Tango which do not occur in W. Murcott. Del2 is a dinucleotide (TC) deletion at 34905517 bp to 34905518 bp (or 34905519 bp to 34905520 bp, due to the nucleotides from 34905517 bp to 34905520 bp on reference genome are TCTC) on scaffold 2. Del3 is a single nucleotide (G) deletion at 1636488 bp associated with a SNP at adjacent position 1636489 bp on scaffold 3, where the deletion is linked with a C while the other haplotype has an A. W. Murcott also had both C allele and A allele at 1636489 bp (-C/GA on Tango and GC/GA on W Murcott). Del4-2 is a single nucleotide (G) deletion at 15692585 bp on scaffold 4, where the other allele and both W. Murcott alleles at the same chromosomal position are G. Del3-3 is a single nucleotide (A) deletion at 45913796 bp on scaffold 3, where the other allele and both W. Murcott alleles at the same chromosomal position are A. Del2-3 is a trinucleotide

(CCT) deletion from 9184855 bp to 9184857 bp on scaffold 2, where the other allele and both W. Murcott alleles had CCT. All of these deletions were confirmed by TA-cloning followed by Sanger sequencing (Table 2.4). Some noise with N's and mismatched nucleotides in Sanger sequencing was evident in some clones, probably due to low sequencing quality (Figure 2.25 - 2.31).

Discussion

For DC-1, DC-7, 3-GP, 5-GP, 33-GP, Del2, Del3, Del4-2, Del3-3 and Del2-3, BLAST results showed the sequences containing the Tango-specific nucleotide or concatenated sequences flanking the short deletion were not present in the reference Clementine genome, indicating those PCR products are less likely from elsewhere in the genome other than the reported or predicted loci.

HRM real-time PCR results for additional Indels showed clear differences of PCR products between W. Murcott and Tango, as expected if the extra PCR product in Tango changed the overall melt curve. For most Indels on derivative melt curves, Tango samples showed extra and minor peaks or had peaks with different melting temperature than W. Murcott. Results for Del3-3 (a single base deletion) did not show significantly different melting temperature or number of peaks in Tango versus W. Murcott, but a slight difference of derivative melting curves between W. Murcott and Tango can still be observed, indicating different products were generated in Tango than W. Murcott. However, it is still likely that there are two different products in Tango for Del3-3, as supported by TA cloning followed by sequencing (Figure 2.30). The similarity in melting curves of Tango and W. Murcott may be due to the candidate marker being a single nucleotide deletion in one of the two alleles in Tango and the sequence context that masked the tiny differences. One of the two NTCs for Del3-3 showed a different curve than the other NTC and might be contaminated.

The real-time PCR results, with different template DNAs than the allele-specific PCR reactions, did not show evidence of the Tango-specific product in W. Murcott samples.

From the sequencing results of TA cloning, DC-1 is a 20 bp heterozygous deletion in Tango which can be located on scaffold 1 from 19024294 bp to 19024313 bp according to *Citrus clementina* genome assembly 1.0 with sequence GCTACTTATAGACTCTTAGT or 19024291 bp to 19024310 bp with sequence AGTGCTACTTATAGACTCTT, since both left and right flanking nucleotides of GCTACTTATAGACTCTT on reference genome are AGT. In addition, the TA cloning results revealed that the deletion and nearby SNPs are linked, indicating the two haplotypes of Tango. The haplotype pattern of the nucleotides within the intact allele was also found in NGS sequencing data of some pummelo varieties at the same region according to JBrowse on Citrus Genome Database (https://www.citrusgenomedb.org). Therefore, this haplotype structure may be from pummelo, an ancestry of W. Murcott and Tango, rather than an alignment error. Other short deletions, DC-7, Del2, Del3, Del4-2, Del3-3 and Del2-3, showed a similar pattern, in which approximately half of Tango clones lacked a few consecutive nucleotides, while the other half of Tango clones contained the sequence, and no W. Murcott clones carried such deletions, indicating the heterozygosity of the short deletions in Tango.

One or more small deletions on one of the two alleles of Tango may be due to errors during repair of a double-strand break that occurred during irradiation. Double strand break repair could affect the genome organization through homologous recombination. Previous studies of DSB induced by a transformed gene expressing a rare cutting restriction endonuclease and following repair in Arabidopsis and tobacco showed that not only could double-strand

break repair induce deletions but also the deletion size varied (KIRIK et al., 2000). In this study, 40% of deletions were accompanied by insertions in tobacco, while no insertions were detected in Arabidopsis, and large deletions were more frequent than small deletions (KIRIK et al., 2000). By identifying the loss of function of a negatively selectable marker gene of the recombined junctions, the average deletion size in tobacco was found to be 1,341 bp, which was larger than 920 bp in Arabidopsis (KIRIK et al., 2000). Therefore, the deletions in Tango may be related to double-strand break repair events. However, if this is the case, then deletion sizes in Tango are much smaller than those observed in Arabidopsis and tobacco. Also, double-strand break repair is error-prone, which could lead to genomic changes such as deletions, insertions or genomic rearrangements (PIPIRAS et al., 1998).

According to the genome browser on Citrus Genome Database, on the Clementine genome, the DC-7 position is within the coding sequence of the gene Ciclev10025094m.g which encodes a hypothetical protein. The characterized gene with most similarity (82% identity of 157 bp) to this gene is the partial coding sequence of Myb-like transcription factor mRNA of *Cocos nucifera*. The 33-GP position is within the intron of the gene Ciclev10025487m.g, whose sequence is most similar to the gene sequence of *Citrus sinensis* nuclear pore complex protein NUP58. The Del2 position is within the coding sequence of the gene Ciclev10014107m.g, which best corresponds to a hypothetical protein in *Citrus clementina,* and a predicted *Citrus sinensis* probable starch synthase 4 with 99% identity. The Del3 position is within an intron of the gene Ciclev10021784m.g, which corresponds to a *Citrus clementina* hypothetical protein. The Del3-3 position is within a gene Ciclev10024146m.g, whose sequence corresponds to a partial coding sequence of a

Citrus clementina hypothetical protein. The other SNPs or Indels were not found within annotated genic regions (Table 2.4). There is no clear evidence suggesting that these SNPs and Indels alter genes that are related to meiosis or cause gamete sterility, but this possibility cannot be excluded.

Advantage of HRM Real-Time PCR

Since most potential Indels were only 1-3 base pairs, conventional PCR with primers flanking the Indels and product separation on agarose gels may not have sufficient resolution to distinguish the PCR products of the two alleles of Tango from those of W. Murcott. HRM real-time PCR has sufficient sensitivity to overcome this limitation and was able to detect differences between PCR products as small as 1 bp. Since the melting temperature of dsDNA depends on the sequence and the occurrence of SNPs or Indels can perturb the melting temperature, heterozygous SNPs or Indels can thus be inferred by comparing melting curves to those of the homozygous counterpart (GIBSON, 2006). Measurement of the melting temperature with high-resolution temperature scanning is the best PCR practice for accurately distinguishing the variant-containing sequences (GIBSON, 2006). By utilizing the double-strand-binding fluorescence, the curves representing the change of signal strength with the change in melting temperature can be used as the profile for a certain dsDNA or dsDNA mixture. The melt curves of real-time PCR products were similar for W. Murcott and Tango, but slight differences still existed due to the two-amplicon mixture in Tango compared to a single amplicon in W. Murcott. PCR grade water

was used as non-template control (NTC) to exclude the possible influence of the secondary structure of primers since fluorescent dye can also bind to primers if they form secondary structures.

Further Sanger Sequencing Verification

Tango is expected to be a heterozygote for novel SNPs/Indels and copy number variants in comparison to W. Murcott, which is homozygous at those loci. The Sanger sequencing results matched the expectation, where there were two genotypes in Tango and one of the two genotypes matches the homozygous genotype in W. Murcott. The frequency of clones that carried the Tango-specific SNPs or Indels was approximately half of the total clones of Tango, indicating Tango is likely heterozygous on these loci in the sampled tissue. However, other possibilities could not be eliminated. Since the samples for PCR and following cloning were leaf samples, most cells of which are mesophyll and derived from cell layer II of shoot apical meristem, it is possible that a minor proportion of cells of leaf were not affected by the mutation (chimera), but their homozygosity was masked by the heterozygous majority. To investigate whether some cells were not affected by the mutation, tissues mainly consisting of cells derived from other layers of the shoot apical meristem need to be investigated.

Indels Previously Discovered in Other Citrus Varieties

In previous studies, 12 Indel polymorphisms were identfied in citrus, ranging from 2 bp to 50 bp, within genes and compared through a diverse citrus population by designing primers that amplified fragments between 150 bp and 350 bp (GARCIA-LOR et al., 2012). These 12 Indel markers could distinguish approximately 60% of genotypes within the whole 90 sample set of the Citrus genus., revealing the interspecific differences between cultivated citrus (GARCIA-LOR et al., 2012). The Indel markers distinguishing Tango from W. Murcott we discovered in our study are throughout whole genome in both genic and intergenic regions, and the sizes are relatively small, ranging from 1 bp to 3 bp. The other two Indels reported by IVIA, DC-1 and DC-7, were validated and the sizes are 20 bp and 13 bp, respectively. Thus, the Indels discovered in Tango are relatively similar in size to those due to natural mutation in citrus germplasm.

Molecular Markers for Other Parent-Mutant Pairs

For grape cultivars, two ALFP markers were found in "Flame Seedless" to differentiate the mutant clone from the parent, with each marker occurring in either only the mutant or only the parent, respectively (SCOTT et al., 2000). The zygosity of AFLP markers is not known because they typically show dominant inheritance of band presence. Pinot Meunier is a periclinal chimera mutated from Pinot Noir, and a difference was found at an SSR locus, where Pinot Meunier had an additional allele besides sharing two alleles with Pinot Moir, which proved to be due to the chimerism of Pinot Meunier (STENKAMP et al., 2009). In

another study on Pinot Noir and its two mutants, Pinot Blanc and Pinot Gris, the combination of several SNP and SSR markers was found to distinguish the three cultivars (VEZZULLI et al., 2012). In addition, cell layer analysis of the three varieties revealed that Pinot Gris is a chimera with different genotypes in cell layers I and II, and a large heterozygous deletion with size over 4 kb on chromosome 2 of cell layer II cells (VEZZULLI et al., 2012). The peach cultivar 'Daguang 1', a natural early-ripening mutant, differed from its parent at one RAPD marker (JIN et al., 1998). In soybean, RAPD markers were developed to differentiate three induced mutants, Coles, Amsoy-71 and 1937 varieties (ATAK et al., 2004). RAPD markers were also developed for identification of *Jatropha curcas* L. mutants developed by gamma radiation (DHAKSHANAMOORTHY et al., 2011). Many of these examples are "sequence anonymous" markers which are less easily interpreted at a molecular level than the markers reported here for Tango.

The Significance of Molecular Markers for Tango

The SNPs and Indels found between Tango and W. Murcott can potentially be developed into molecular markers that distinguish the two varieties in a precise, quick and inexpensive method Accurate criteria are necessary to classify cultivars that are morphologically identical, otherwise it would be difficult to define a legal framework for cultivar discrimination (STAUB AND MEGLIC, 1993). Since Tango mandarin is patented, the distinctness from W. Murcott mandarin in genetic diversity should be described unambiguously in measurable terms. The easiest way to distinguishing the two varieties is

to count the seeds of fruit produced in an open pollination condition. However, the fulfillment of this method is time-consuming and may not be possible outside the fruiting season. Furthermore, if W. Murcott fruits are produced in isolation from other pollen sources, either by distance or by netting to exclude pollinators, the fruit of both cultivars will have very few seeds. The developed molecular markers can significantly help reduce the time and labor needed to distinguish the two varieties, and should be applicable at any development stage of plants. In addition, phenotypic differences are subject to modification by the environment, whereas molecular markers provide effective, robust and measurable traits, so that the protected patent can in turn support incentive for continued investments by institutions.

Besides patent protection, to trace the genomic changes during the propagation generations of Tango, the developed markers can also be applied to each propagation generation to investigate if those SNPs and Indels are associated with initial gamma irradiation or subsequent natural mutation. Also since gamma irradiation could affect only one or a few cells in the shoot apical meristem, the following cell division may result in the chimerism of Tango, with some cells carry the SNPs and Indels while others do not. To better investigate which parts of the meristem were mutated, the SNPs and Indels which are associated with gamma irradiation can be genotyped on tissues derived from different cell layers of the shoot apical meristem.

References

Agarwal, M., N. Shrivastava, and H. Padh, 2008. Advances in molecular marker techniques and their applications in plant sciences. *Plant Cell Reports* 27:617-631.

Akkurt, M., A. Cakir, M. Shidfar, F. Mutaf, and G. Soylemezoglu, 2013. Using seedlessness-related molecular markers in grapevine breeding for seedlessness via marker-assisted selection into Muscat of Hamburg x Sultani progeny. *Turkish Journal of Biology* 37:101-105.

Amar, M.H., M.K. Biswas, Z.W. Zhang, and W.W. Guo, 2011. Exploitation of SSR, SRAP and CAPS-SNP markers for genetic diversity of Citrus germplasm collection. *Scientia Horticulturae* 128:220-227.

Atak, C., S. Alikamanoğlu, L. Açık, and Y. Canbolat, 2004. Induced of plastid mutations in soybean plant (Glycine max L. Merrill) with gamma radiation and determination with RAPD. *Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis* 556:35-44.

Barkley, N.A., R.R. Krueger, C.T. Federici, and M.L. Roose, 2009. What phylogeny and gene genealogy analyses reveal about homoplasy in citrus microsatellite alleles. *Plant Systematics and Evolution* 282:71-86.

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

Bernet, G., M. Bretó, and M. Asins, 2004. Expressed sequence enrichment for candidate gene analysis of citrus tristeza virus resistance. *Theoretical and Applied Genetics* 108:592-602.

Bhattramakki, D., M. Dolan, M. Hanafey, R. Wineland, D. Vaske, J.C. Register, S.V. Tingey, and A. Rafalski, 2002. Insertion-deletion polymorphisms in 3 ' regions of maize genes occur frequently and can be used as highly informative genetic markers. *Plant Molecular Biology* 48:539-547.

Britten, R.J., L. Rowen, J. Williams, and R.A. Cameron, 2003. Majority of divergence between closely related DNA samples is due to indels. *Proceedings of the National Academy of Sciences of the United States of America* 100:4661-4665.

Butelli, E., C. Licciardello, Y. Zhang, J. Liu, S. Mackay, P. Bailey, G. Reforgiato-Recupero, and C. Martin, 2012. Retrotransposons control fruit-specific, cold-dependent accumulation of anthocyanins in blood oranges. *The Plant Cell* 24:1242-1255.

Chen, C.X., K.D. Bowman, Y.A. Choi, P.M. Dang, M.N. Rao, S. Huang, J.R. Soneji, T.G. McCollum, and F.G. Gmitter, 2008. EST-SSR genetic maps for Citrus sinensis and Poncirus trifoliata. *Tree Genetics & Genomes* 4:1-10.

Cheng, F.S. and M.L. Roose, 1995. Origin and inheritance of dwarfing by the Citrus rootstock Poncirus trifoliata 'Flying Dragon'. *Journal of the American Society for Horticultural Science* 120:286-291.

Close, T., S. Wanamaker, M. Lyon, G. Mei, C. Davies, and M. Roose, 2006. A GeneChip® for citrus. *Plant & Animal Genome XIV Conference*:82.

Crowley, J.R., 2011. A molecular genetic approach to evaluate a novel seedless phenotype found in Tango, a new variety of mandarin developed from gamma-irradiated W. Murcott. Ph.D. Dissertation, University of California, Riverside.

Cuenca, J., P. Aleza, A. Garcia-Lor, P. Ollitrault, and L. Navarro, 2016. Fine mapping for identification of Citrus alternaria brown spot candidate resistance genes and development of new SNP markers for marker-assisted selection. *Frontiers in plant science* 7:1948.

Cuenca, J., P. Aleza, A. Vicent, D. Brunel, P. Ollitrault, and L. Navarro, 2013. Genetically based location from triploid populations and gene ontology of a 3.3-mb genome region linked to Alternaria brown spot resistance in citrus reveal clusters of resistance genes. *PLoS One* 8:e76755.

Das, S., H.D. Upadhyaya, R. Srivastava, D. Bajaj, C. Gowda, S. Sharma, S. Singh, A.K. Tyagi, and S.K. Parida, 2015. Genome-wide insertion-deletion (InDel) marker discovery and genotyping for genomics-assisted breeding applications in chickpea. *DNA Research* 22:377-386.

Davey, J.W., P.A. Hohenlohe, P.D. Etter, J.Q. Boone, J.M. Catchen, and M.L. Blaxter, 2011. Genome-wide genetic marker discovery and genotyping using next-generation sequencing. *Nature Reviews Genetics* 12:499-510.

Davis, M.W. and M. Hammarlund, 2006. Single-nucleotide polymorphism mapping. *C. elegans: Methods and Applications* 351:75-92.

Dhakshanamoorthy, D., R. Selvaraj, and A.L.A. Chidambaram, 2011. Induced mutagenesis in Jatropha curcas L. using gamma rays and detection of DNA polymorphism through RAPD marker. *Comptes Rendus Biologies* 334:24-30.

Drenkard, E., B.G. Richter, S. Rozen, L.M. Stutius, N.A. Angell, M. Mindrinos, R.J. Cho, P.J. Oefner, R.W. Davis, and F.M. Ausubel, 2000. A simple procedure for the analysis of single nucleotide polymorphisms facilitates map-based cloning in Arabidopsis. *Plant physiology* 124:1483-1492.

Edwards, M.D., T. Helentjaris, S. Wright, and C.W. Stuber, 1992. Molecular-marker-facilitated investigations of quantitative trait loci in maize : 4. Analysis based on

genome saturation with isozyme and restriction fragment length polymorphism markers. *Theoretical and Applied Genetics* 83:765-774.

Fang, D., C. Federici, and M. Roose, 1997. Development of molecular markers linked to a gene controlling fruit acidity in citrus. *Genome* 40:841-849.

Federici, C.T., D.Q. Fang, R.W. Scora, and M.L. Roose, 1998. Phylogenetic relationships within the genus Citrus (Rutaceae) and related genera as revealed by RFLP and RAPD analysis. *Theoretical and Applied Genetics* 96:812-822.

Fujii, H., T. Shimada, K. Nonaka, M. Kita, T. Kuniga, T. Endo, Y. Ikoma, and M. Omura, 2013. High-throughput genotyping in citrus accessions using an SNP genotyping array. *Tree Genetics & Genomes* 9:145-153.

Gambetta, G., A. Gravina, C. Fasiolo, C. Fornero, S. Galiger, C. Inzaurralde, and F. Rey, 2013. Self-incompatibility, parthenocarpy and reduction of seed presence in 'Afourer' mandarin. *Scientia Horticulturae* 164:183-188.

Garcia-Lor, A., F. Luro, L. Navarro, and P. Ollitrault, 2012. Comparative use of InDel and SSR markers in deciphering the interspecific structure of cultivated citrus genetic diversity: a perspective for genetic association studies. *Molecular Genetics and Genomics* 287:77-94.

Garritano, S., F. Gemignani, C. Voegele, T. Nguyen-Dumont, F. Le Calvez-Kelm, D. De Silva, F. Lesueur, S. Landi, and S.V. Tavtigian, 2009. Determining the effectiveness of High Resolution Melting analysis for SNP genotyping and mutation scanning at the TP53 locus. *BMC Genetics* 10:1.

Gaudet, M., A.G. Fara, I. Beritognolo, and M. Sabatti, 2009. Allele-specific PCR in SNP genotyping. *Single Nucleotide Polymorphisms: Methods and Protocols* 578:415-424.

Gibson, N.J., 2006. The use of real-time PCR methods in DNA sequence variation analysis. *Clinica Chimica Acta* 363:32-47.

Gmitter Jr, F.G., C. Chen, M.N. Rao, and J.R. Soneji, 2007. Citrus fruits, p. 265-279, Fruits and Nuts. Springer.

Gulsen, O., A. Uzun, I. Canan, U. Seday, and E. Canihos, 2010. A new citrus linkage map based on SRAP, SSR, ISSR, POGP, RGA and RAPD markers. *Euphytica* 173:265-277.

Guo, F., H.W. Yu, Z. Tang, X.L. Jiang, L. Wang, X. Wang, Q. Xu, and X.X. Deng, 2015. Construction of a SNP-based high-density genetic map for pummelo using RAD sequencing. *Tree Genetics & Genomes* 11:2.

Hayashi, K., N. Hashimoto, M. Daigen, and I. Ashikawa, 2004. Development of PCR-based SNP markers for rice blast resistance genes at the Piz locus. *Theoretical and Applied Genetics* 108:1212-1220.

Hayashi, K., H. Yoshida, and I. Ashikawa, 2006. Development of PCR-based allele-specific and InDel marker sets for nine rice blast resistance genes. *Theoretical and Applied Genetics* 113:251-260.

Huang, J., H.B. Luo, W. Wei, and Y.P. Hou, 2014. A novel method for the analysis of 20 multi-Indel polymorphisms and its forensic application. *Electrophoresis* 35:487-493.

Jiang, D., Q.L. Ye, F.S. Wang, and L. Cao, 2010. The mining of citrus EST-SNP and its application in cultivar discrimination. *Agricultural Sciences in China* 9:179-190.

Jin, Y., Y. Zhang, D. Chen, and S. Zhang, 1998. Identification of peach early-ripening mutant 'Daguang 1' by RAPD markers and cloning of specific fragment. *Journal of fruit science* 2:103-106.

Kepiro, J. and M. Roose, 2010. AFLP markers closely linked to a major gene essential for nucellar embryony (apomixis) in Citrus maxima × Poncirus trifoliata. *Tree genetics & genomes* 6:1-11.

Kirik, A., S. Salomon, and H. Puchta, 2000. Species-specific double-strand break repair and genome evolution in plants. *EMBO Journal* 19:5562-5566.

Kumar, P., V.K. Gupta, A.K. Misra, D.R. Modi, and B.K. Pandey, 2009. Potential of molecular markers in plant biotechnology. *Plant Omics* 2:141-162.

Li, Z.Q., T.M. Li, Y.J. Wang, and Y. Xu, 2015. Breeding new seedless grapes using in ovulo embryo rescue and marker-assisted selection. *In Vitro Cellular & Developmental Biology-Plant* 51:241-248.

McCaskill, A. and J. Giovannoni, 2002. Use of molecular markers for fruit crop improvement, p. 283-297, Molecular Techniques in Crop Improvement. Springer.

Miles, C., G. Sargent, G. Phear, and M. Meuth, 1990. DNA-sequence analysis of gamma-radiation - induced deletions and insertions at the Aprt locus of hamster-cells. *Molecular Carcinogenesis* 3:233-242.

Mills, R.E., C.T. Luttig, C.E. Larkins, A. Beauchamp, C. Tsui, W.S. Pittard, and S.E. Devine, 2006. An initial map of insertion and deletion (INDEL) variation in the human genome. *Genome Research* 16:1182-1190.

Moore, G.A., 2001. Oranges and lemons: clues to the taxonomy of Citrus from molecular markers. *TRENDS in Genetics* 17:536-540.

Moose, S.P. and R.H. Mumm, 2008. Molecular plant breeding as the foundation for 21st century crop improvement. *Plant Physiology* 147:969-977.

Novelli, V.M., M.A. Takita, and M.A. Machado, 2004. Identification and analysis of single nucleotide polymorphisms (SNPs) in citrus. *Euphytica* 138:227-237.

Ollitrault, P., Y. Froelicher, D. Dambier, F. Luro, and M. Yamamoto, 2007. Seedlessness and ploidy manipulation. *Citrus Genetics, Breeding and Biotechnology*:197-218.

Ollitrault, P., A. Garcia-Lor, J. Terol, F. Curk, F. Ollitrault, M. Talón, and L. Navarro, 2014. Comparative values of SSRs, SNPs and InDels for citrus genetic diversity analysis. *Acta Horticulturae*.

Ollitrault, P., J. Terol, C. Chen, C.T. Federici, S. Lotfy, I. Hippolyte, F. Ollitrault, A. Berard, A. Chauveau, and J. Cuenca, 2015. Comparative genetic mapping between clementine, pummelo and sweet orange and the interspecific structure of the clementine genome. *Xii International Citrus Congress-International Society of Citriculture* 1065:561-573.

Ollitrault, P., J. Terol, C. Chen, C.T. Federici, S. Lotfy, I. Hippolyte, F. Ollitrault, A. Berard, A. Chauveau, J. Cuenca, G. Costantino, Y. Kacar, L. Mu, A. Garcia-Lor, Y. Froelicher, P. Aleza, A. Boland, C. Billot, L. Navarro, F. Luro, M.L. Roose, F.G. Gmitter, M. Talon, and D. Brunel, 2012. A reference genetic map of C. clementina hort. ex Tan.; citrus evolution inferences from comparative mapping. *BMC Genomics* 13:593.

Paterson, A.H., E.S. Lander, J.D. Hewitt, S. Peterson, S.E. Lincoln, and S.D. Tanksley, 1988. Resolution of quantitative traits into mendelian factors by using a complete linkage map of restriction fragment length polymorphisms. *Nature* 335:721-726.

Pipiras, E., A. Coquelle, A. Bieth, and M. Debatisse, 1998. Interstitial deletions and intrachromosomal amplification initiated from a double-strand break targeted to a mammalian chromosome. *EMBO Journal* 17:325-333.

Raman, H., R. Raman, R. Wood, and P. Martin, 2006. Repetitive indel markers within the ALMT1 gene conditioning aluminium tolerance in wheat (Triticum aestivum L.). *Molecular Breeding* 18:171-183.

Roose, M.L. and T.E. Williams, 2007. Mandarin tree named 'Tango'. U.S. Patent No. PP17,863.

Ruiz, C. and M.J. Asins, 2003. Comparison between Poncirus and Citrus genetic linkage maps. *Theoretical and Applied Genetics* 106:826-836.

Ruiz, C., M.P. Breto, and M.J. Asins, 2000. A quick methodology to identify sexual seedlings in citrus breeding programs using SSR markers. *Euphytica* 112:89-94.

Sankaranarayanan, K., R. Taleei, S. Rahmanian, and H. Nikjoo, 2013. Ionizing radiation and genetic risks. XVII. Formation mechanisms underlying naturally occurring DNA deletions in the human genome and their potential relevance for bridging the gap between induced DNA double-strand breaks and deletions in irradiated germ cells. *Mutation Research/Reviews in Mutation Research* 753:114-130.

Scott, K.D., E.M. Ablett, L.S. Lee, and R.J. Henry, 2000. AFLP markers distinguishing an early mutant of Flame Seedless grape. *Euphytica* 113:245-249.

Semagn, K., A. Bjornstad, and M.N. Ndjiondjop, 2006. An overview of molecular marker methods for plants. *African Journal of Biotechnology* 5:2540-2568.

Shiokai, S., H. Kitashiba, K. Shirasawa, K. Nagano, and T. Nishio, 2009. Leaf-punch method to prepare a large number of PCR templates from plants for SNP analysis. *Molecular Breeding* 23:329-336.

Singh, B.D. and A.K. Singh, 2015. Hybridization-based markers, p. 19-46, Marker-Assisted Plant Breeding: Principles and Practices. Springer India, New Delhi.

Staub, J. and V. Meglic, 1993. Molecular genetic markers and their legal relevance for cultivar discrimination: a case study in cucumber. *HortTechnology* 3:291-300.

Steele, K.A., R. Ogden, R. McEwing, H. Briggs, and J. Gorham, 2008. InDel markers distinguish Basmatis from other fragrant rice varieties. *Field Crops Research* 105:81-87.

Stenkamp, S.H.G., M.S. Becker, B.H.E. Hill, R. Blaich, and A. Forneck, 2009. Clonal variation and stability assay of chimeric Pinot Meunier (Vitis vinifera L.) and descending sports. *Euphytica* 165:197-209.

Sun, H., O.R. Lee, Y.J. Kim, S.K. Jeong, J.G. In, W.S. Kwon, S.Y. Kim, and D.C. Yang, 2010. Identification of 'Chunpoong' among Panax ginseng cultivars using real time PCR and SNP marker. *Journal of Ginseng Research* 34:47-50.

Väli, Ü., M. Brandström, M. Johansson, and H. Ellegren, 2008. Insertion-deletion polymorphisms (indels) as genetic markers in natural populations. *BMC genetics* 9:8.

Vasemagi, A., R. Gross, D. Palm, T. Paaver, and C.R. Primmer, 2010. Discovery and application of insertion-deletion (INDEL) polymorphisms for QTL mapping of early life-history traits in Atlantic salmon. *BMC Genomics* 11:1.

Vezzulli, S., L. Leonardelli, U. Malossini, M. Stefanini, R. Velasco, and C. Moser, 2012. Pinot blanc and Pinot gris arose as independent somatic mutations of Pinot noir. *Journal of Experimental Botany* 63:6359-6369.

Vignal, A., D. Milan, M. SanCristobal, and A. Eggen, 2002. A review on SNP and other types of molecular markers and their use in animal genetics. *Genetics Selection Evolution* 34:275-305.

Wang, Y.Q., S.X. Li, X.T. Zhang, Y.J. Wang, and C.H. Zhang, 2016. Isolation and analysis of differentially expressed genes during ovule abortion in the seedless grape. *Scientia Horticulturae* 211:376-383.

Wijerathna, Y., 2015. Marker assisted selection: biotechnology tool for rice molecular breeding. *Advances in Crop Science and Technology*:1-4.

Wu, G.A., S. Prochnik, J. Jenkins, J. Salse, U. Hellsten, F. Murat, X. Perrier, M. Ruiz, S. Scalabrin, J. Terol, M.A. Takita, K. Labadie, J. Poulain, A. Couloux, K. Jabbari, F. Cattonaro, C. Del Fabbro, S. Pinosio, A. Zuccolo, J. Chapman, J. Grimwood, F.R. Tadeo, L.H. Estornell, J.V. Munoz-Sanz, V. Ibanez, A. Herrero-Ortega, P. Aleza, J. Perez-Perez, D. Ramon, D. Brunel, F. Luro, C. Chen, W.G. Farmerie, B. Desany, C. Kodira, M. Mohiuddin, T. Harkins, K. Fredrikson, P. Burns, A. Lomsadze, M. Borodovsky, G. Reforgiato, J. Freitas-Astua, F. Quetier, L. Navarro, M. Roose, P. Wincker, J. Schmutz, M. Morgante, M.A. Machado, M. Talon, O. Jaillon, P. Ollitrault, F. Gmitter, and D. Rokhsar, 2014. Sequencing of diverse mandarin, pummelo and orange genomes reveals complex history of admixture during citrus domestication. *Nature Biotechnology* 32:656-662.

Xu, Q., L.L. Chen, X. Ruan, D. Chen, A. Zhu, C. Chen, D. Bertrand, W.B. Jiao, B.H. Hao, M.P. Lyon, J. Chen, S. Gao, F. Xing, H. Lan, J.W. Chang, X. Ge, Y. Lei, Q. Hu, Y. Miao, L. Wang, S. Xiao, M.K. Biswas, W. Zeng, F. Guo, H. Cao, X. Yang, X.W. Xu, Y.J. Cheng, J. Xu, J.H. Liu, O.J. Luo, Z. Tang, W.W. Guo, H. Kuang, H.Y. Zhang, M.L. Roose, N. Nagarajan, X.X. Deng, and Y. Ruan, 2013. The draft genome of sweet orange (Citrus sinensis). *Nature Genetics* 45:59-66.

Yamaki, S., H. Ohyanagi, M. Yamasaki, M. Eiguchi, T. Miyabayashi, T. Kubo, N. Kurata, and K.I. Nonomura, 2013. Development of INDEL markers to discriminate all genome types rapidly in the genus Oryza. *Breeding Science* 63:246-254.

Yu, Y., C. Chen, and F.G. Gmitter, 2016. QTL mapping of mandarin (Citrus reticulata) fruit characters using high-throughput SNP markers. *Tree Genetics & Genomes* 12:77.

Table 2.1 Summary of Allele-Specific Primers Used to Analyze SNPs and Indels in Tango and W. Murcott

SNP/Indel/Control	Primer name	Sequence	Control SSR
DC-1	DC1_deletion_ASPCR_F	GGGTATGCTCATAACAGTAT	CF-ATG05
	DC1-R	CGTCACGAGAAGTCATTGCT	
DC-7	DC7_deletion_ASPCR_F	CATAAAACTAATTCTACGAT	CF-ATG05
	DC7-R	GATAGGTGTGTGTGTTTGAGTTGC	
3-GP	3GP_ASPCR_F	TCATTCTTGTCAATCTTTCT	CF-ATG05
	3-GP-R	CGAGGGATCCTCTCTTGATG	
5-GP	5GP_ASPCR_F	TTCATTCTCGACTATGTTGC	CF-ATG05
	5-GP-R	CGATCGAGTCCAAACACCTC	
33-GP	33-GP-F	CGGAACTACAGGTGCTCCAT	CF-AT01
	33-GP_ASPCR_R	GACTAAGCTTCATCTCAAAA	
CF-ATG05	SSR#133 forward	GGTTCTTTGTTTGGGTGCTG	
	SSR#133 reverse	ATGAACTCGAACACCACCTTTC	
CF-AT01	SSR#131 forward	TGTGAGTGTTTGTGCGTGTG	
	SSR#131 reverse	GCCTCCGGGAATGTTAGATA	

Table 2.2 Summary of Real-Time PCR Primers Used to Analyze Indels in Tango and W. Murcott

Indel	Primer name	Sequence
DC-1	YZ_DC1_5F_SYBR	AAAGAACTAGGTTGGGACCTAAA
	YZ_DC1_6R_SYBR	CCTCAAACAAACATCCCTCATC
DC-7	UCR56	ACTGGGTCAGCACCCAAA
	UCR57	ACGGTGAAGGACAATTTAAGGAAAG
Del2	UCR27	ACTTAAACGTCCCGTTGCT
	UCR28	GTACCCGAAACTACGCTGTC
Del3	UCR29	TCCATTTCTGCTTCCTGTAGATAA
	UCR30	CTGGACTATAGCAGCAAGTCTAC
Del4-2	UCR35	GACGTCATGAATTACCTCTCAGAT
	UCR36	TACAAGGGCAAGATATGGCAA
Del3-3	UCR48	AATCGATCCTCAAGACCTAATGTT
	UCR49	AACGGCGTCGAAAGTGATAG
Del2-3	UCR80	CGGTGCCAAGTTGTTGTC
	UCR81	ACATCATTCAGGGCTCACA

Table 2.3 Summary of Primers Used to Amplify Segments for TA-Cloning for Candidate Markers that Differentiate Tango and W. Murcott

Indel	Primer name	Sequence
DC-1	DC1-F	GTTCTGGAGCACCGTTGAAT
	DC1-R	CGTCACGAGAAGTCATTGCT
DC-7	UCR132	CCAACTCAAGAACTACTA
	UCR133	GGAGAGGAAAGCTAAAT
Del2	UCR134	CCATTCGATCAGGATA
	UCR135	AAATCTGCATCATTTGG
Del3	UCR136	GAGGTACAAGATGAGAG
	UCR137	CACTCTTTCTCCACATA
Del4-2	UCR138	TCTTGCTACTTGTGATG
	UCR139	GAAAGAGCTACAAGTTAAG
Del3-3	UCR140	CTAACAAACTCACATAGTC
	UCR141	GAGGACTTACGAATCAA
Del2-3	UCR142	CTGATATCTCCCATCTTG
	UCR143	CAACTCATCTCTAGGTTC

Table 2.4 Summary of SNPs and Indels

SNP/Indel name	Genomic position on the reference genome	Alleles in W. Murcott	Alleles in Tango	Annotation *
DC-1	Scaffold 1: 19024291 bp - 19024313 bp	AGTGCTACTTATAGACTCTTAGT/ AGTGCTACTTATAGACTCTTAGT	AGTGCTACTTATAGACTCTTAGT/ AGT	
DC-7	Scaffold 7: 6988443 bp - 6988455 bp	TTGGTGAAGAAAA/ TTGGTGAAGAAAA	TTGGTGAAGAAAA/-	Ciclev10025094m.g
3-GP	Scaffold 1: 6204314 bp	C/C	C/T	
5-GP	Scaffold 1: 19505366 bp	G/G	G/C	
33-GP	Scaffold 7: 2717333 bp	A/A	A/T	Ciclev10025487m.g
Del2	Scaffold 2: 34905517 bp - 34905520 bp	TCTC/TCTC	TCTC/TC	Ciclev10014107m.g
Del3	Scaffold 3: 1636488 bp - 1636489 bp	GA/GC	GA/C	Ciclev10021784m.g
Del4-2	Scaffold 4: 15692585 bp	C/C	C/-	
Del3-3	Scaffold 3: 45913796 bp	A/A	A/-	Ciclev10024146m.g
Del2-3	Scaffold 2: 9184855 bp - 9184857 bp	CCT/CCT	CCT/-	

* Annotation for Clementine genome on Citrus Genome Database (www.citrusgenomedb.org)

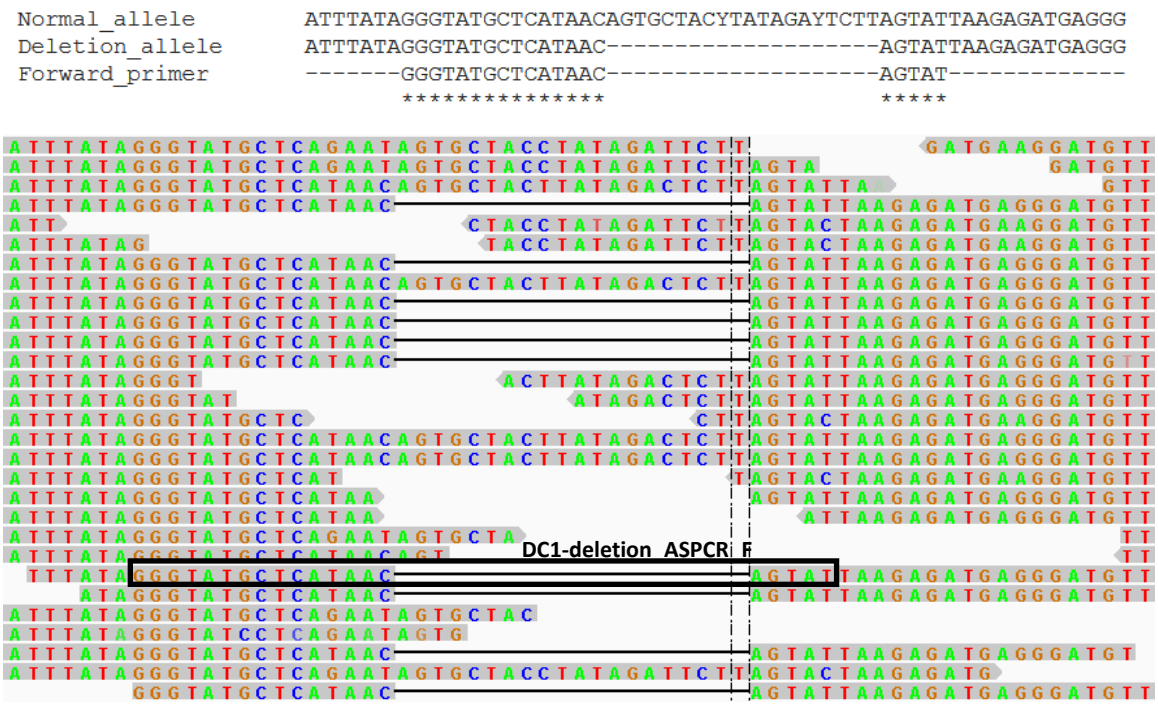


Figure 2.1 DC-1 Short Deletion Forward Primer Design. Illumina reads from Tango leaf DNA are aligned to *Citrus clementina* reference genome v1.0. Figure shows the region spanning the deleted region of DC-1 (black lines) and deletion-specific primer position (black box). The sequence of DC1_deletion_ASPCR_F primer GGGTATGCTCATAACAGTAT is the join of GGGTATGCTCATAAC and AGTAT.

Figure 2.2 DC-7 Short Deletion Forward Primer Design. Illumina reads from Tango leaf DNA are aligned to *Citrus clementina* reference genome v1.0. Figure shows the region spanning the deleted region of DC-7 (black lines) and deletion-specific primer position (black box). The sequence of DC7_deletion_ASPCR_F primer CATAAAACTAATTCTACGAT is the join of CATAAAACTAATTCTACG and AT.

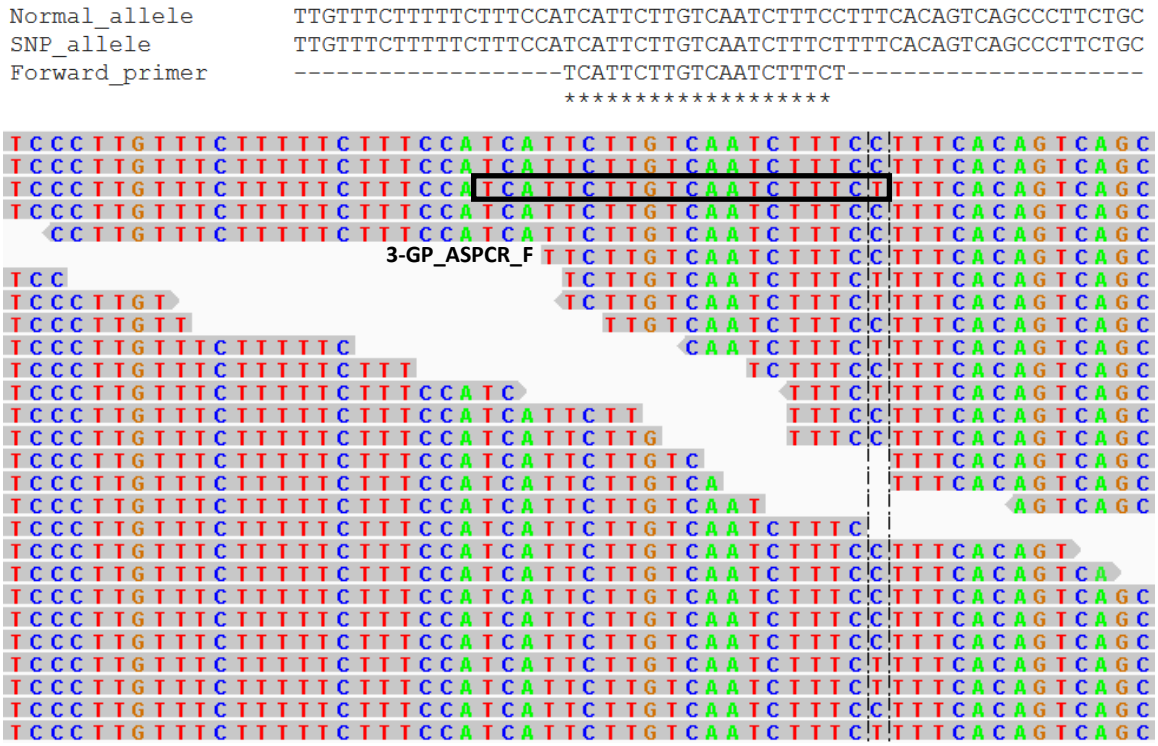


Figure 2.3 3-GP SNP Forward Primer Design. Illumina reads from Tango leaf DNA are aligned to *Citrus clementina* reference genome v1.0. Figure shows the region spanning the SNP position of 3-GP (C/T in dashed box) and SNP-specific primer position (black box). The SNP position is contained at the 3'-end of the sequence of 3GP_ASPCR_F primer TCATTCTTGTCAATCTTTCT, where T is a Tango-specific nucleotide in this locus.

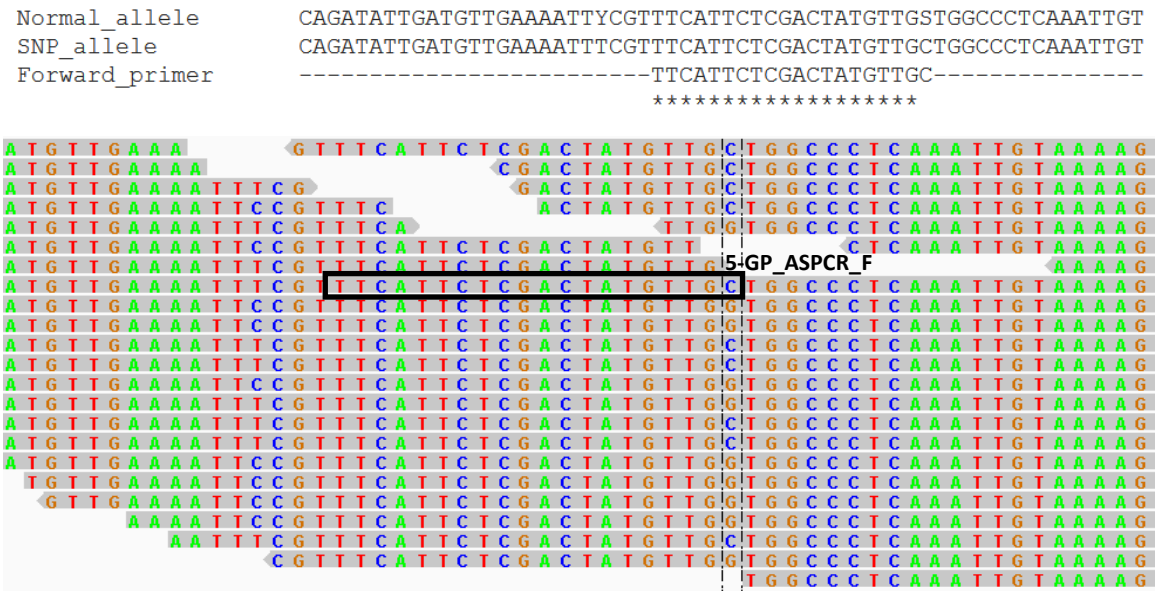


Figure 2.4 5-GP SNP Forward Primer Design. Illumina reads from Tango leaf DNA are aligned to *Citrus clementina* reference genome v1.0. Figure shows the region spanning the SNP position of 5-GP (C/G in dashed box) and SNP-specific primer position (black box). The SNP position is contained at the 3'-end of the sequence of 5GP_ASPCR_F primer TTCATTCTCGACTATGTTGC, where C is a Tango-specific nucleotide in this locus.

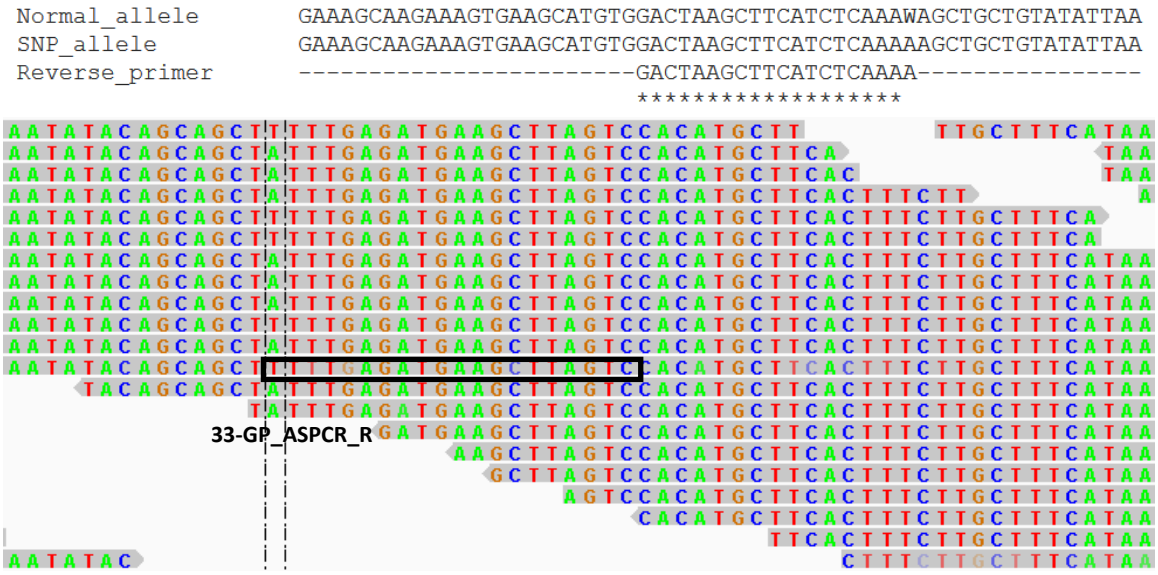


Figure 2.5 33-GP SNP Reverse Primer Design. Illumina reads from Tango leaf DNA are aligned to *Citrus clementina* reference genome v1.0. Figure shows the region spanning the SNP position of 33-GP (T/A in dashed box) and SNP-specific primer position (black box). The SNP position is contained at the 3'-end of the sequence of 33-GP_ASPCR_R primer GACTAAGCTTCATCTCAAAA, where A is a Tango-specific nucleotide in this locus.

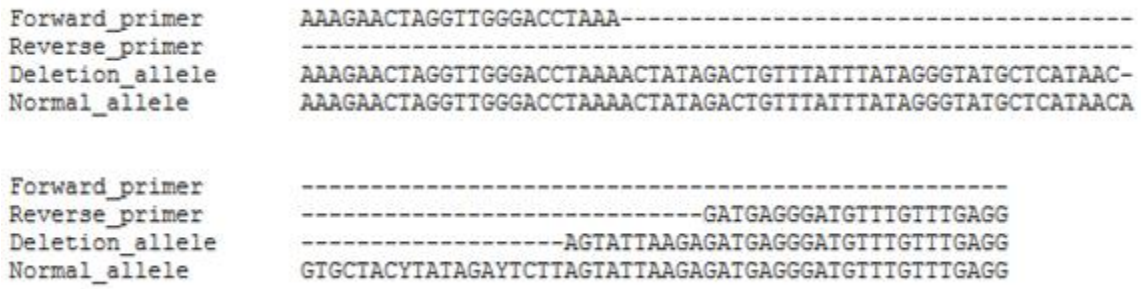


Figure 2.6 DC-1 Real-Time PCR Primer Design. Sequences of normal and deletion alleles of Tango mandarin showing alignment of forward primer (YZ_DC1_5F_SYBR) and reverse primer (YZ_DC1_6R_SYBR) to Scaffold_1:19024232-19024342.

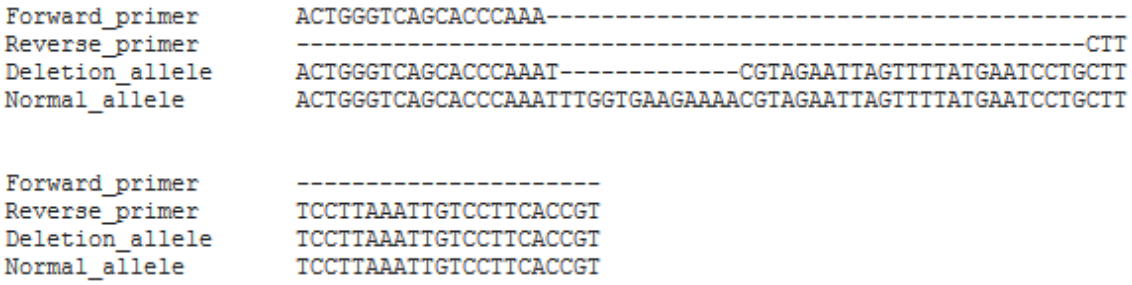


Figure 2.7 DC-7 Real-Time PCR Primer Design. Sequences of normal and deletion alleles of Tango mandarin showing alignment of forward primer (UCR56) and reverse primer (UCR57) to Scaffold_7:6988424-6988505.

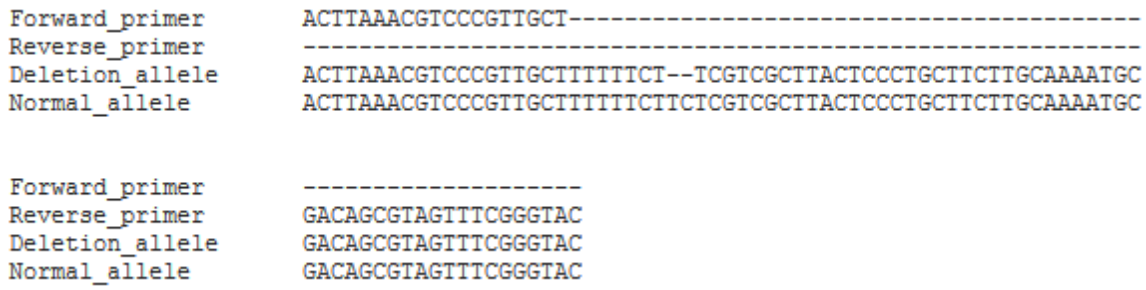


Figure 2.8 Del2 Real-Time PCR Primer Design. Sequences of normal and deletion alleles of Tango mandarin showing alignment of forward primer (UCR27) and reverse primer (UCR28) to Scaffold_2:34905491-34905570.

```
Normal_allele       TCCATTTCTGCTTCCTGTAGATAACTGTCTCTAACATACCCAAGCGMATTCCTTTGTGAC
Deletion_allele     TCCATTTCTGCTTCCTGTAGATAACTGTCTCTAACATACCCAAGC-MATTCCTTTGTGAC
Forward_primer      TCCATTTCTGCTTCCTGTAGATAA----------------------------------
Reverse_primer      ------------------------------------------------------------


Normal_allele       ATAAAAATGTGGGTAGACTTGCTGCTATAGTCCAG
Deletion_allele     ATAAAAATGTGGGTAGACTTGCTGCTATAGTCCAG
Forward_primer      ---------------------------------
Reverse_primer      ------------GTAGACTTGCTGCTATAGTCCAG
```

Figure 2.9 Del3 Real-Time PCR Primer Design. Sequences of normal and deletion alleles of Tango mandarin showing alignment of forward primer (UCR29) and reverse primer (UCR30) to Scaffold_3:1636444-1636538.

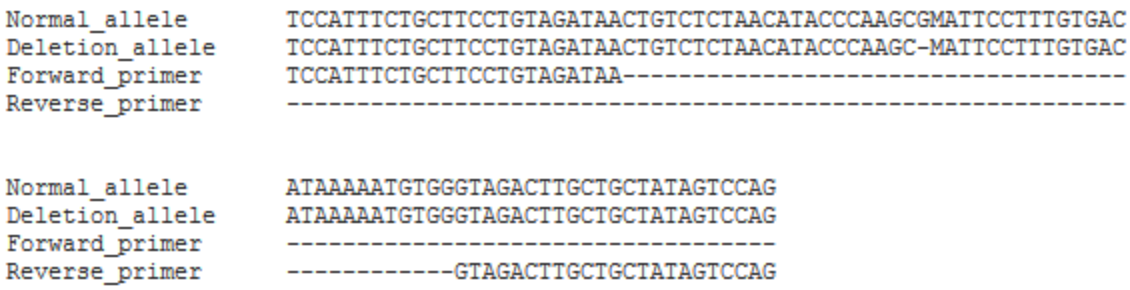


Figure 2.10 Del4-2 Real-Time PCR Primer Design. Sequences of normal and deletion alleles of Tango mandarin showing alignment of forward primer (UCR35) and reverse primer (UCR36) to Scaffold_4:15692537-15692629.

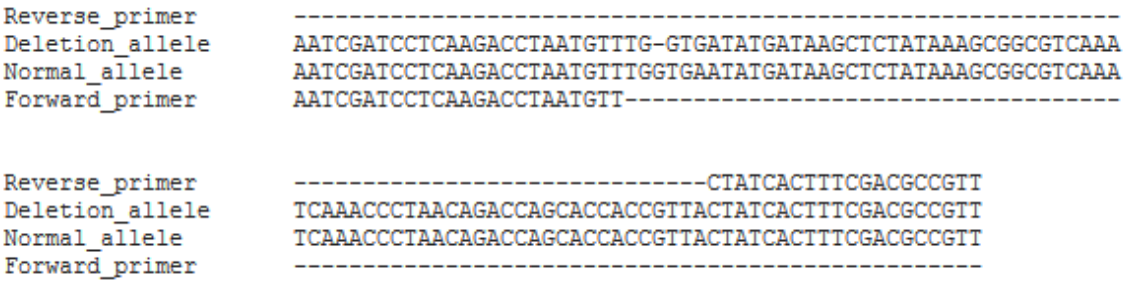


Figure 2.11 Del3-3 Real-Time PCR Primer Design. Sequences of normal and deletion alleles of Tango mandarin showing alignment of forward primer (UCR48) and reverse primer (UCR49) to Scaffold_3:45913767-45913876.

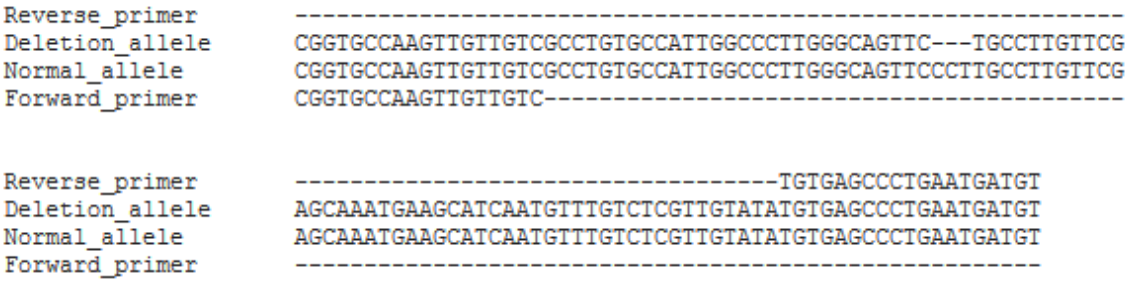


Figure 2.12 Del2-3 Real-Time PCR Primer Design. Sequences of normal and deletion alleles of Tango mandarin showing alignment of forward primer (UCR81) and reverse primer (UCR82) to Scaffold_2:9184810-9184923.

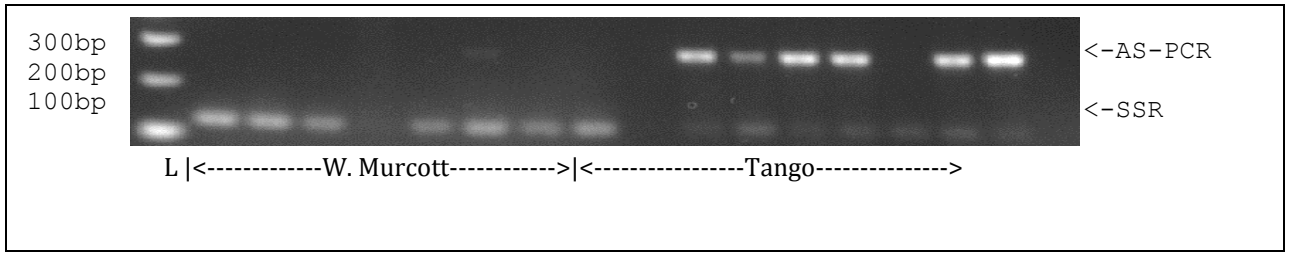


Figure 2.13 Allele-Specific PCR for DC-1 Short Deletion. Eight W. Murcott leaf DNA samples and eight Tango leaf DNA samples amplified with DC-1 short deletion primer sets (DC1_deletion_ASPCR_F and DC1-R) and background control (SSR#133 CF-ATG05) primer sets (SSR#133 forward and SSR#133 reverse).

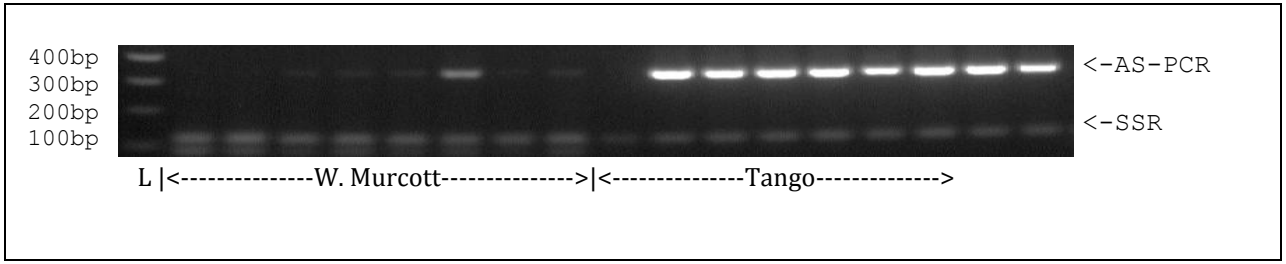


Figure 2.14 Allele-Specific PCR for DC-7 Short Deletion. Nine W. Murcott leaf DNA samples and eight Tango leaf DNA samples amplified with DC-7 short deletion primer sets (DC7_deletion_ASPCR_F and DC7-R) and background control (SSR#133 CF-ATG05) primer sets (SSR#133 forward and SSR#133 reverse).

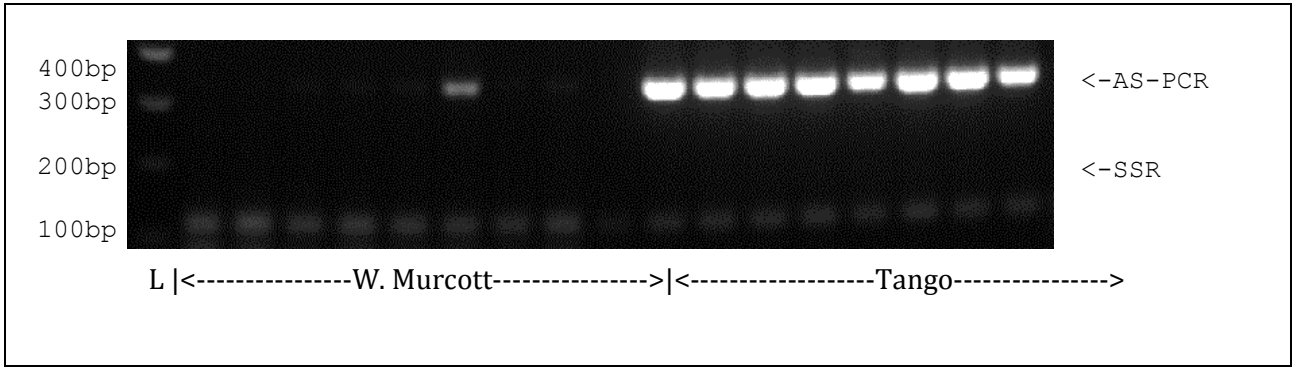


Figure 2.15 Allele-Specific PCR for 3-GP SNP. Nine W. Murcott leaf DNA samples and eight Tango leaf DNA samples amplified with 3-GP SNP primer sets (3GP_ASPCR_F and 3GP_R) and background control (SSR#133 CF-ATG05) primer sets (SSR#133 forward and SSR#133 reverse).

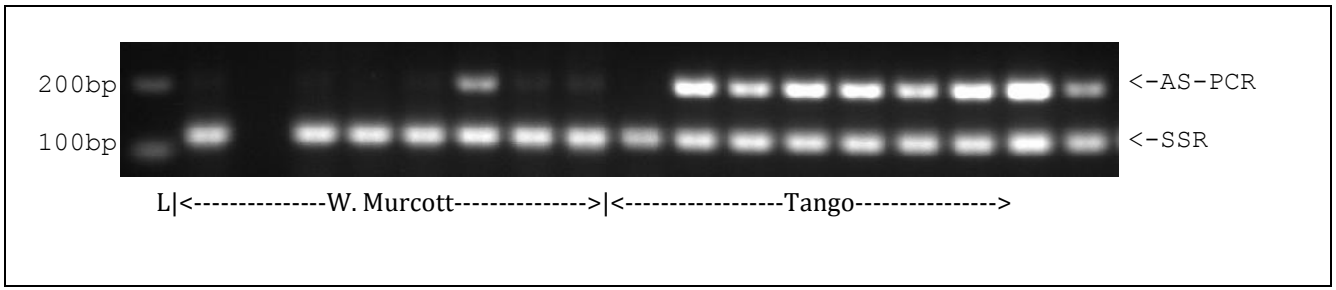


Figure 2.16 Allele-Specific PCR for 5-GP SNP. Nine W. Murcott leaf DNA samples and eight Tango leaf DNA samples amplified with 3-GP SNP primer sets (5GP_ASPCR_F and 5GP_R) and background control (SSR#133 CF-ATG05) primer sets (SSR#133 forward and SSR#133 reverse).

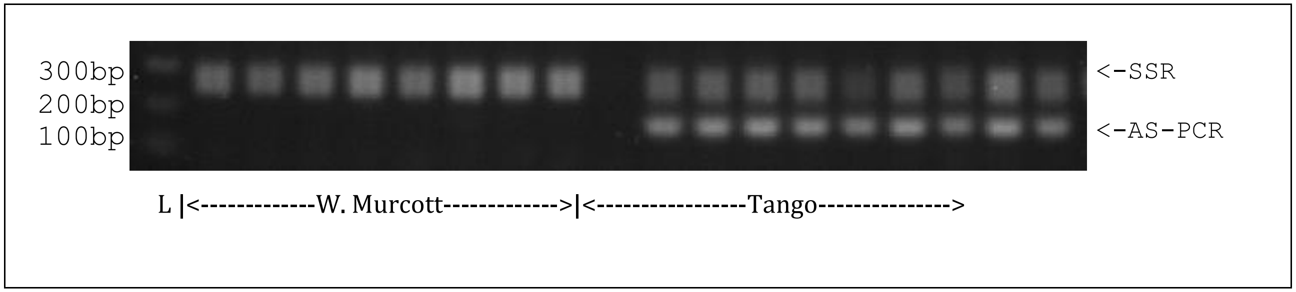


Figure 2.17 Allele-Specific PCR for 33-GP SNP. Nine W. Murcott leaf DNA samples and nine Tango leaf DNA samples amplified with 33-GP SNP primer sets (33-GP_F and 33-GP_ASPCR _R) and background control (SSR#131 CF-AT01) primer sets (SSR#131 forward and SSR#131 reverse).

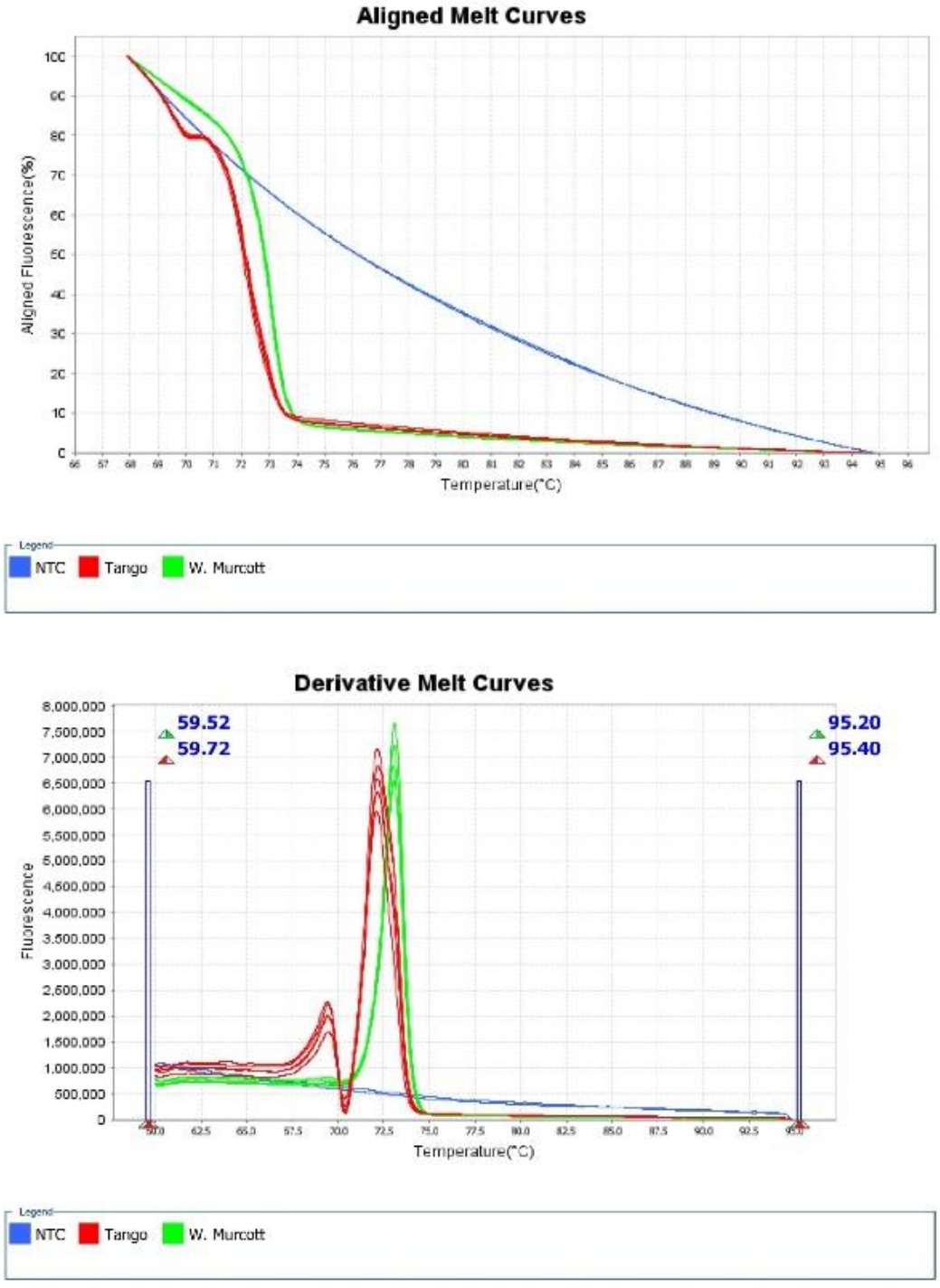


Figure 2.18 Real-Time PCR of DC-1 Short Deletion. Aligned melt curves (upper panel) and derivative melt curves (lower panel) of five W. Murcott leaf DNA samples (green) and five Tango leaf DNA samples (red) with NTC control (blue).

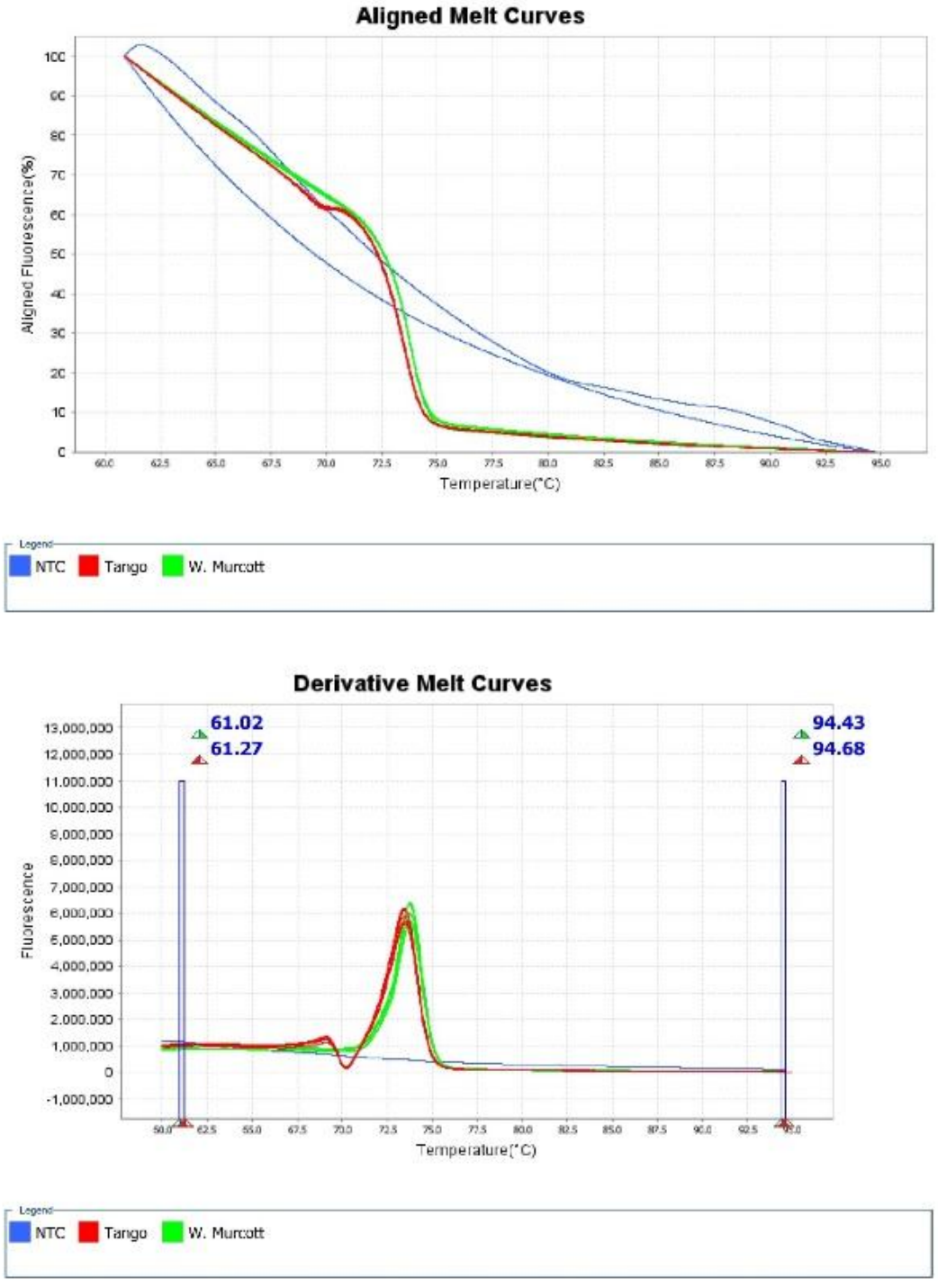


Figure 2.19 Real-Time PCR of DC-7 Short Deletion. Aligned melt curves (upper panel) and derivative melt curves (lower panel) of five W. Murcott leaf DNA samples (green) and five Tango leaf DNA samples (red) with NTC control (blue).

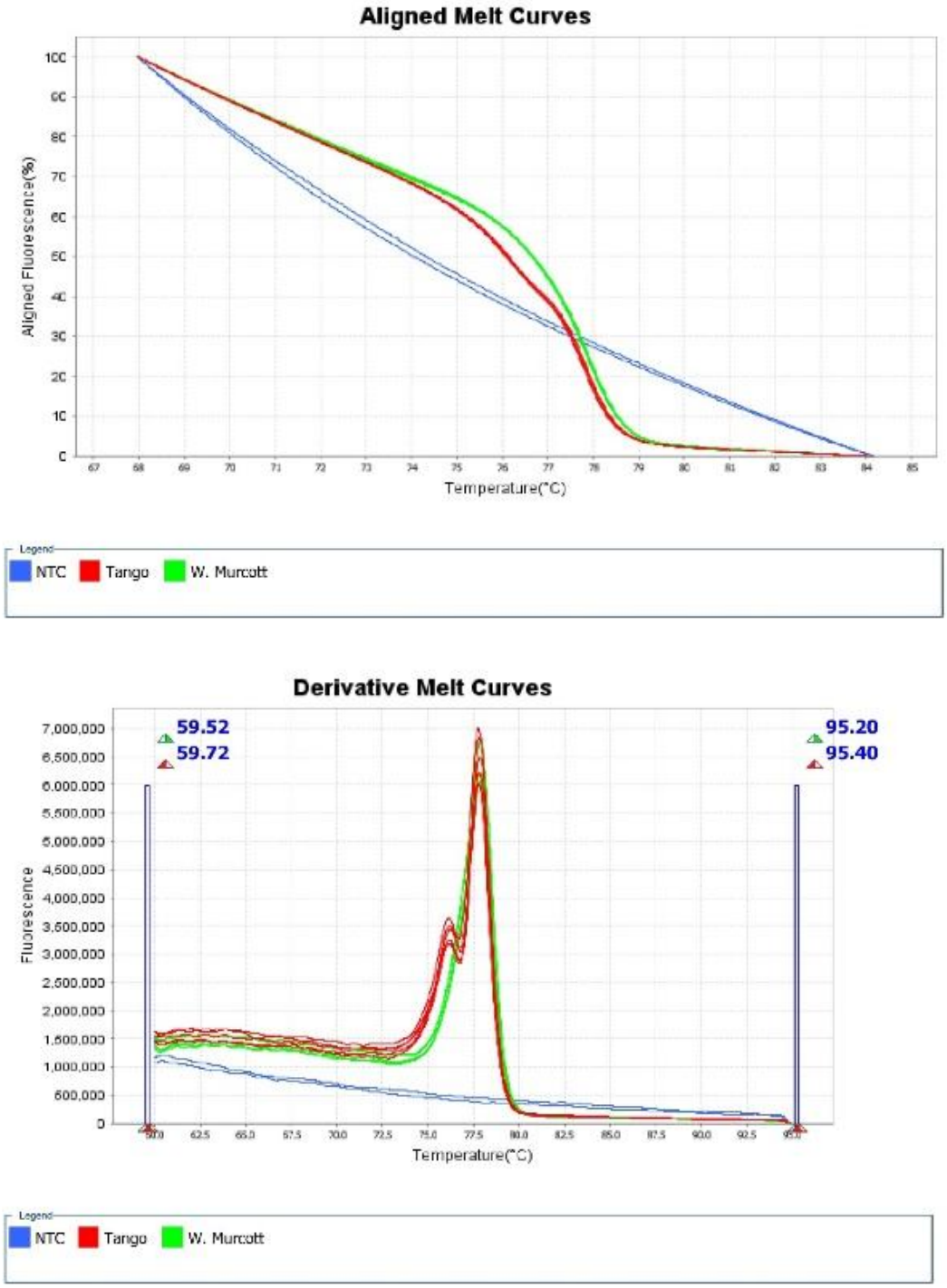


Figure 2.20 Real-Time PCR of Del2 Short Deletion. Aligned melt curves (upper panel) and derivative melt curves (lower panel) of five W. Murcott leaf DNA samples (green) and five Tango leaf DNA samples (red) with NTC control (blue).

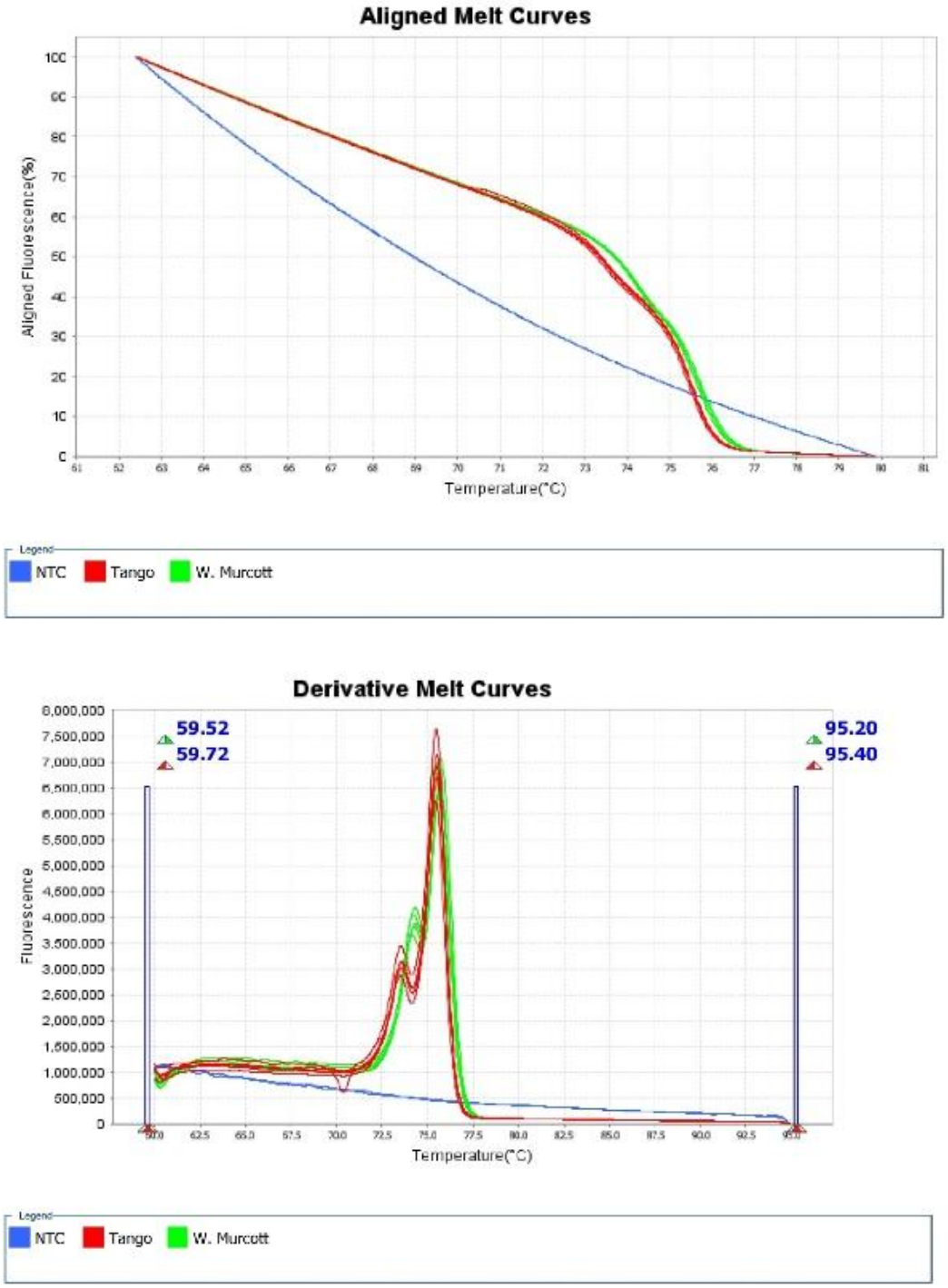


Figure 2.21 Real-Time PCR of Del3 Short Deletion. Aligned melt curves (upper panel) and derivative melt curves (lower panel) of five W. Murcott leaf DNA samples (green) and five Tango leaf DNA samples (red) with NTC control (blue).

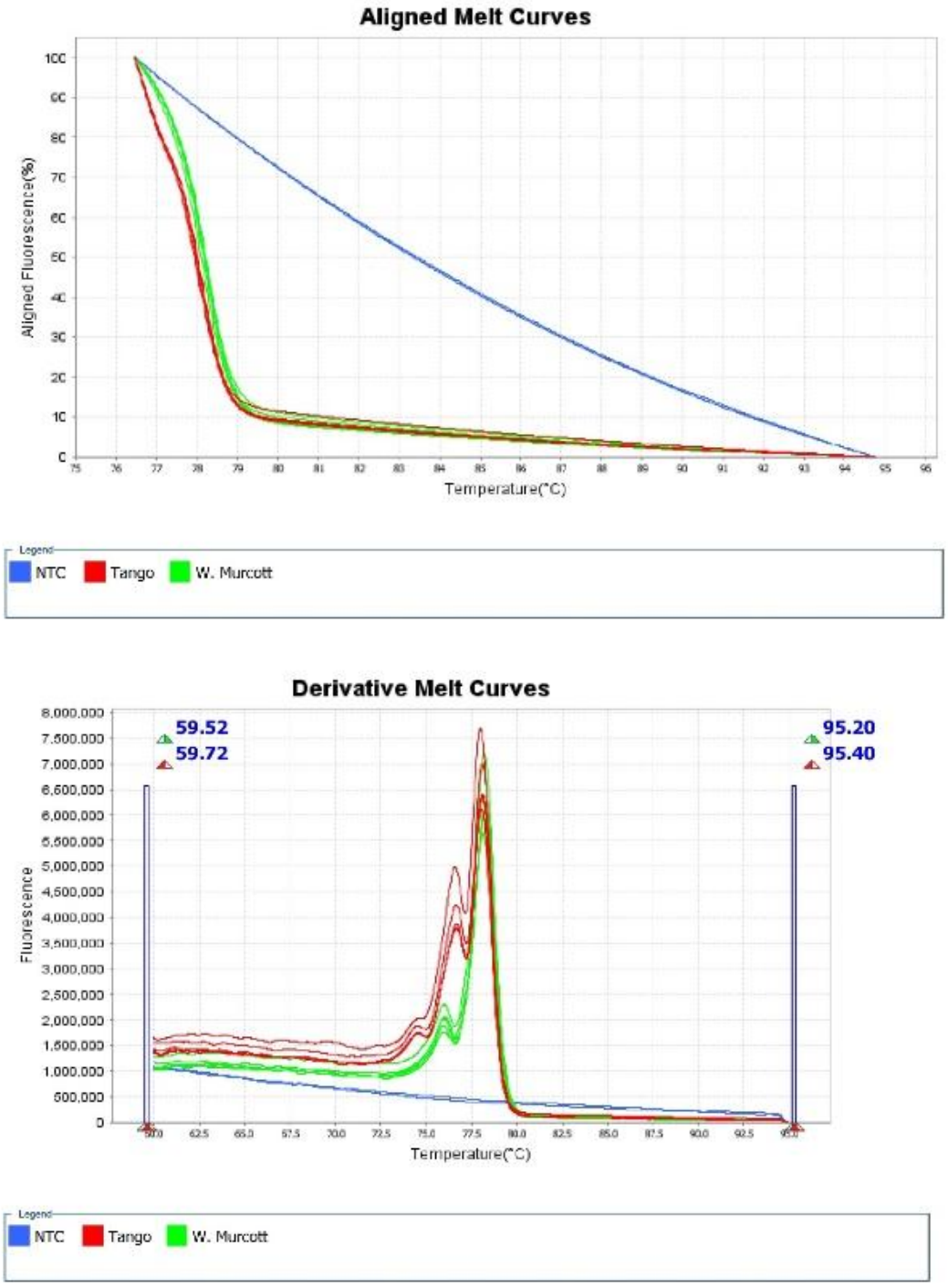


Figure 2.22 Real-Time PCR of Del4-2 Short Deletion. Aligned melt curves (upper panel) and derivative melt curves (lower panel) of five W. Murcott leaf DNA samples (green) and five Tango leaf DNA samples (red) with NTC control (blue).

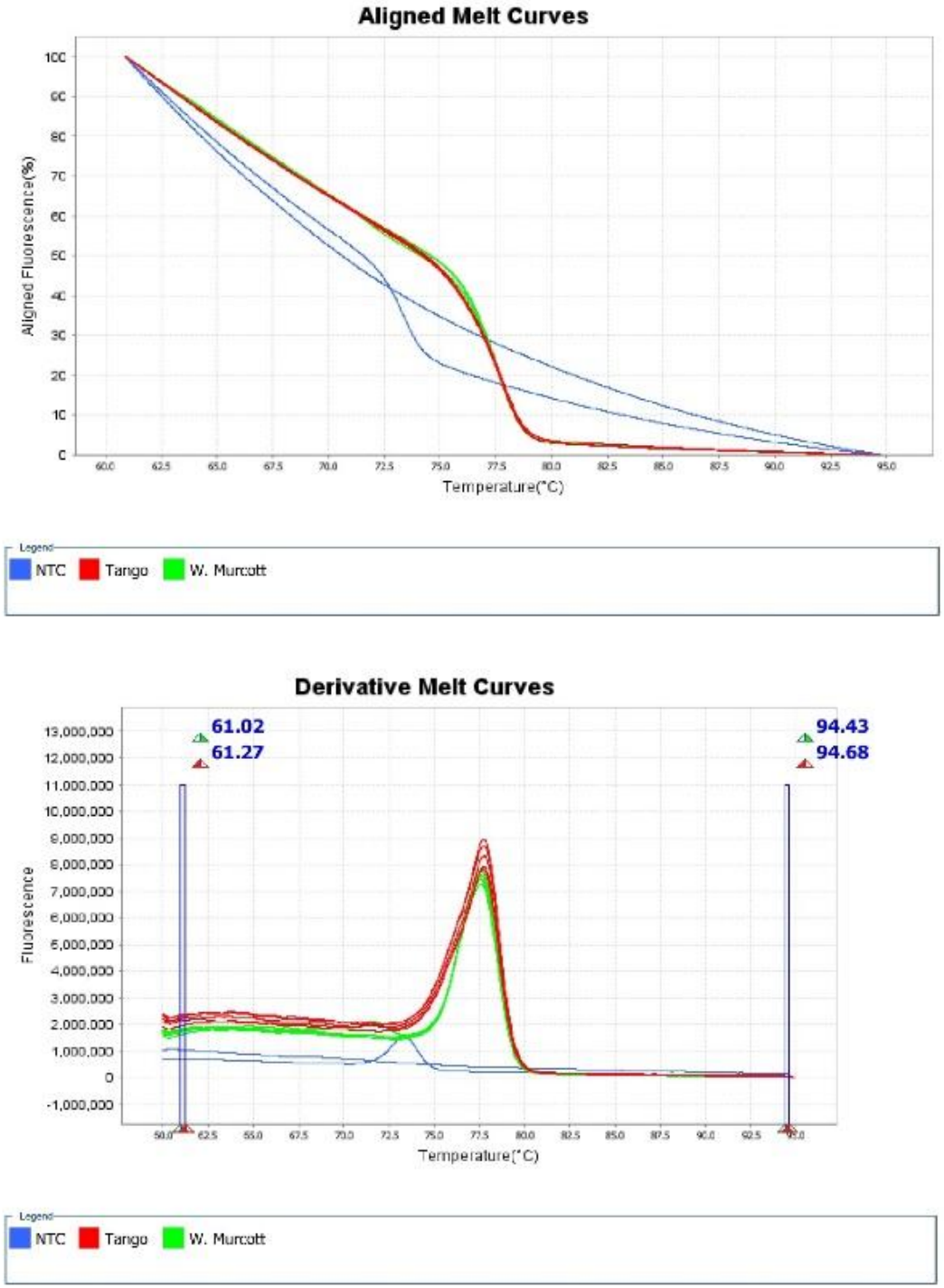


Figure 2.23 Real-Time PCR of Del3-3 Short Deletion. Aligned melt curves (upper panel) and derivative melt curves (lower panel) of five W. Murcott leaf DNA samples (green) and five Tango leaf DNA samples (red) with NTC control (blue).

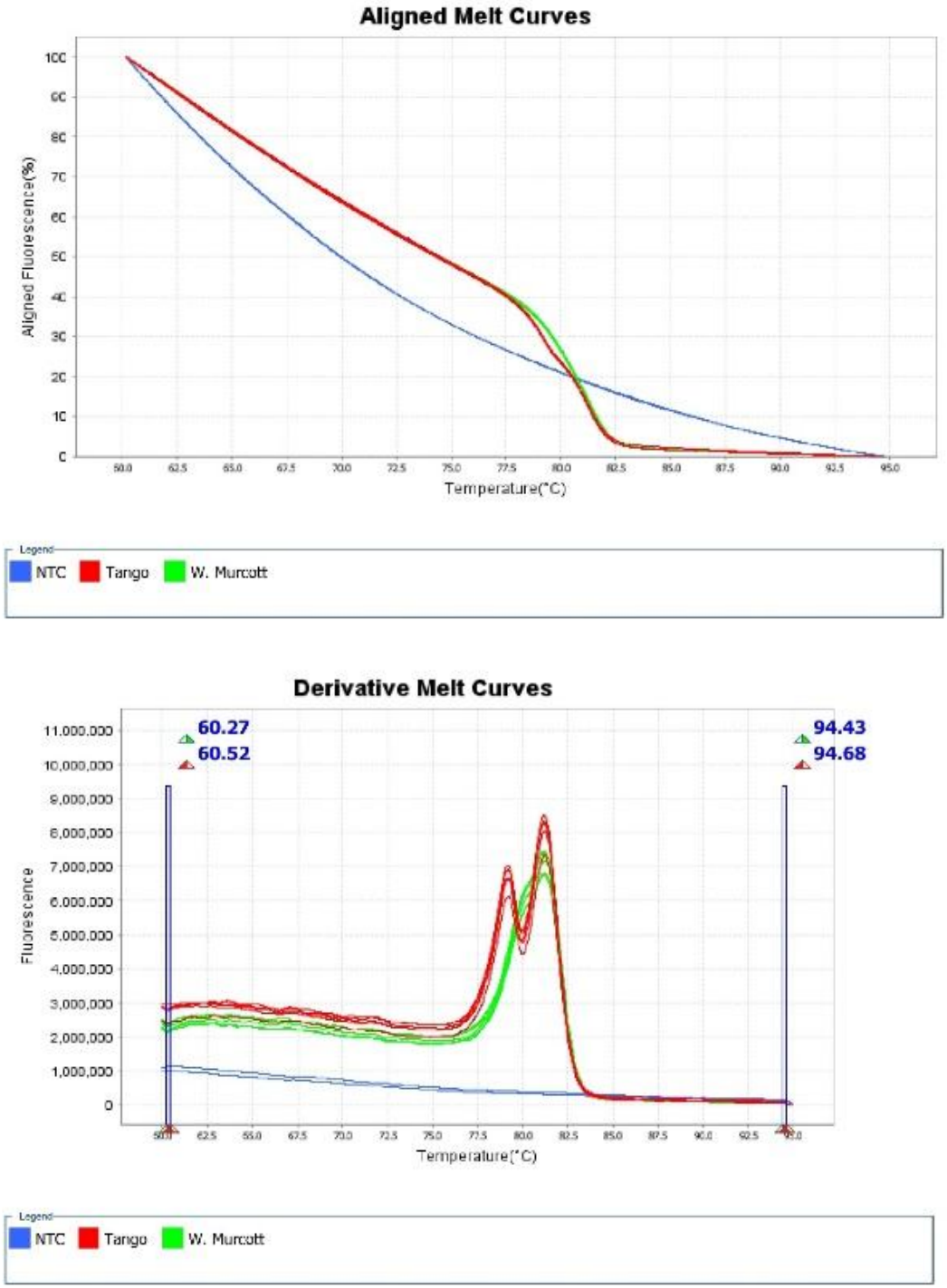


Figure 2.24 Real-Time PCR of Del2-3 Short Deletion. Aligned melt curves (upper panel) and derivative melt curves (lower panel) of five W. Murcott leaf DNA samples (green) and five Tango leaf DNA samples (red) with NTC control (blue).

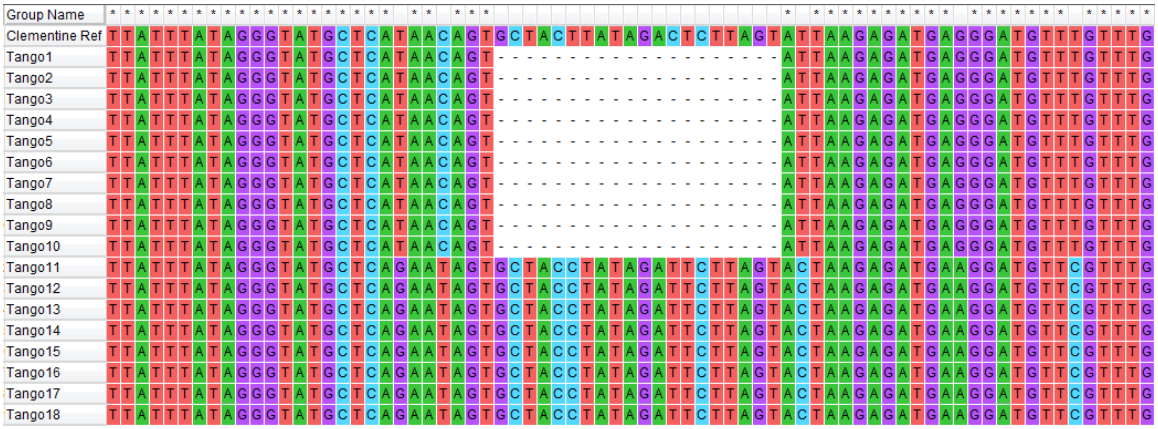


Figure 2.25 Alignment of Partial Sequences Including DC-1 Indel from Sequencing Result of TA-Cloning. 18 sequences of Tango clones (sequence rows named "Tango" in the first column) aligned to *Citrus clementina* reference genome (first sequence row named "Clementine Ref"). Sequences of W. Murcott clones are not shown.

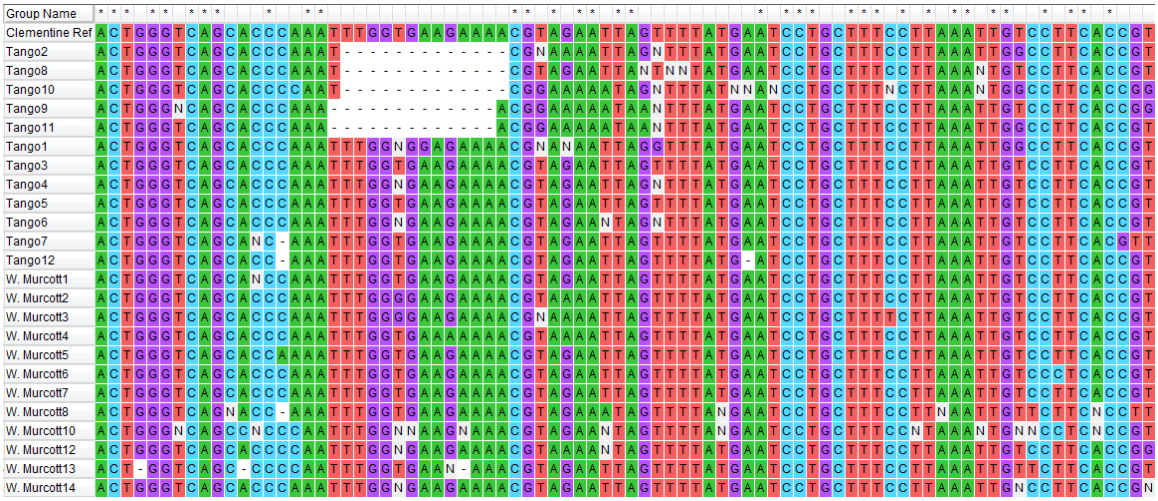


Figure 2.26 Alignment of Partial Sequences Including DC-7 Indel from Sequencing Result of TA-Cloning. 12 sequences of Tango clones (sequence rows named "Tango" in the first column) and 12 sequences of W. Murcott clones (sequence rows named "W. Murcott" in the first column) aligned to *Citrus clementina* reference genome (first sequence row named "Clementine Ref").

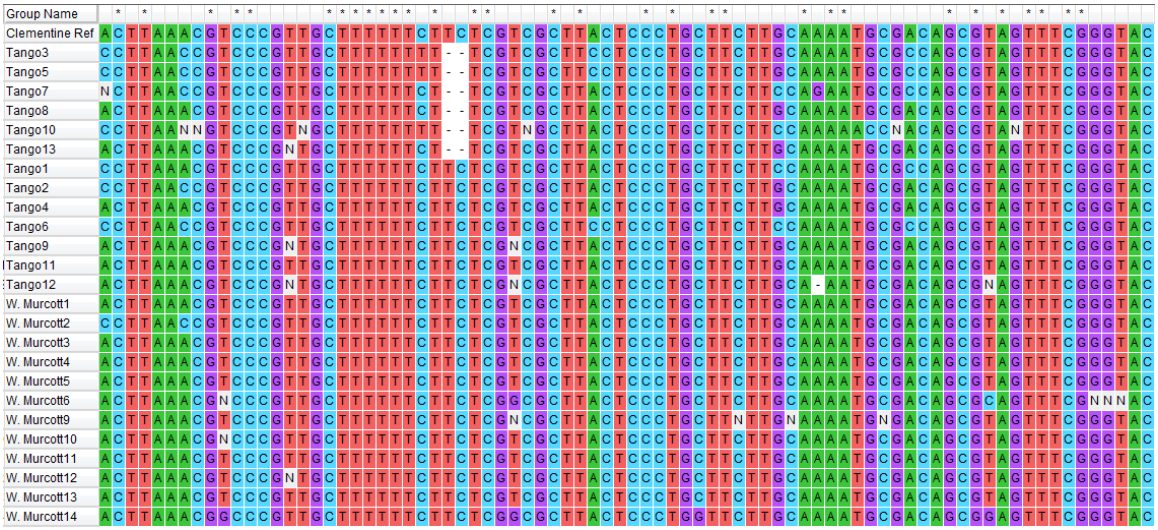


Figure 2.27 Alignment of Partial Sequences Including Del2 Indel from Sequencing Result of TA-Cloning. 13 sequences of Tango clones (sequence rows named "Tango" in the first column) and 12 sequences of W. Murcott clones (sequence rows named "W. Murcott" in the first column) aligned to *Citrus clementina* reference genome (first sequence row named "Clementine Ref").

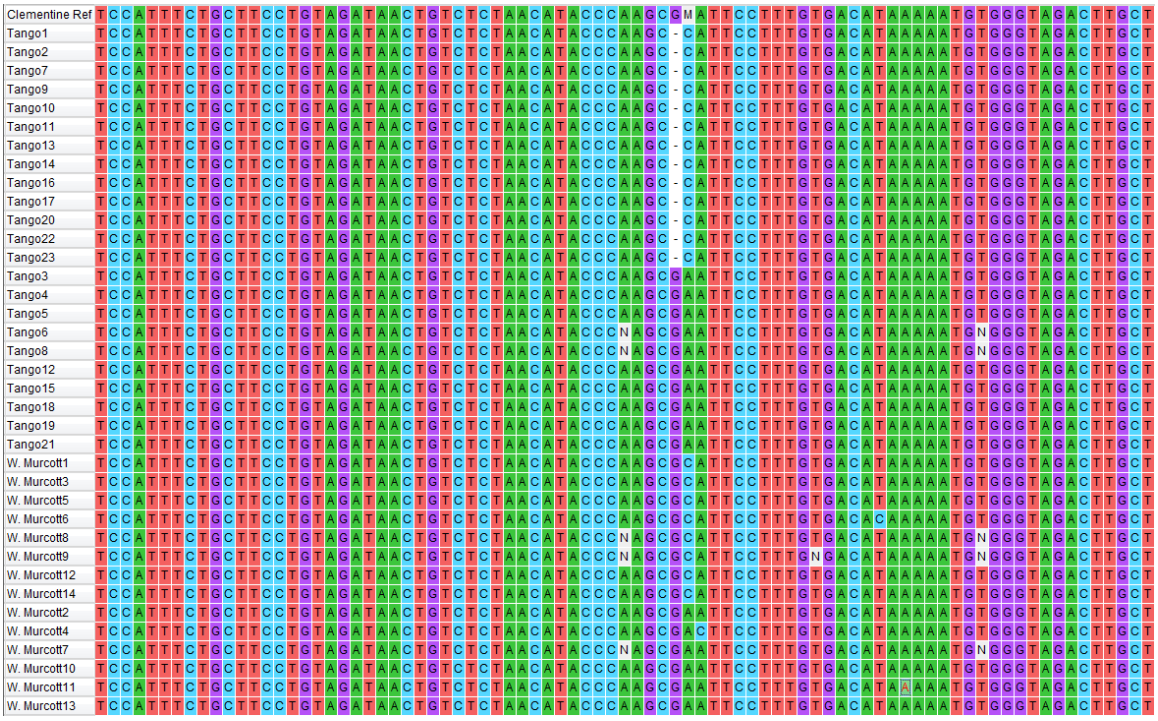


Figure 2.28 Alignment of Partial Sequences Including Del3 Indel from Sequencing Result of TA-Cloning. 23 sequences of Tango clones (sequence rows named "Tango" in the first column) and 14 sequences of W. Murcott clones (sequence rows named "W. Murcott" in the first column) aligned to *Citrus clementina* reference genome (first sequence row named "Clementine Ref").

Figure 2.29 Alignment of Partial Sequences Including Del4-2 Indel from Sequencing Result of TA-Cloning. Six sequences of Tango clones (sequence rows named "Tango" in the first column) and 17 sequences of W. Murcott clones (sequence rows named "W. Murcott" in the first column) aligned to *Citrus clementina* reference genome (first sequence row named "Clementine Ref").

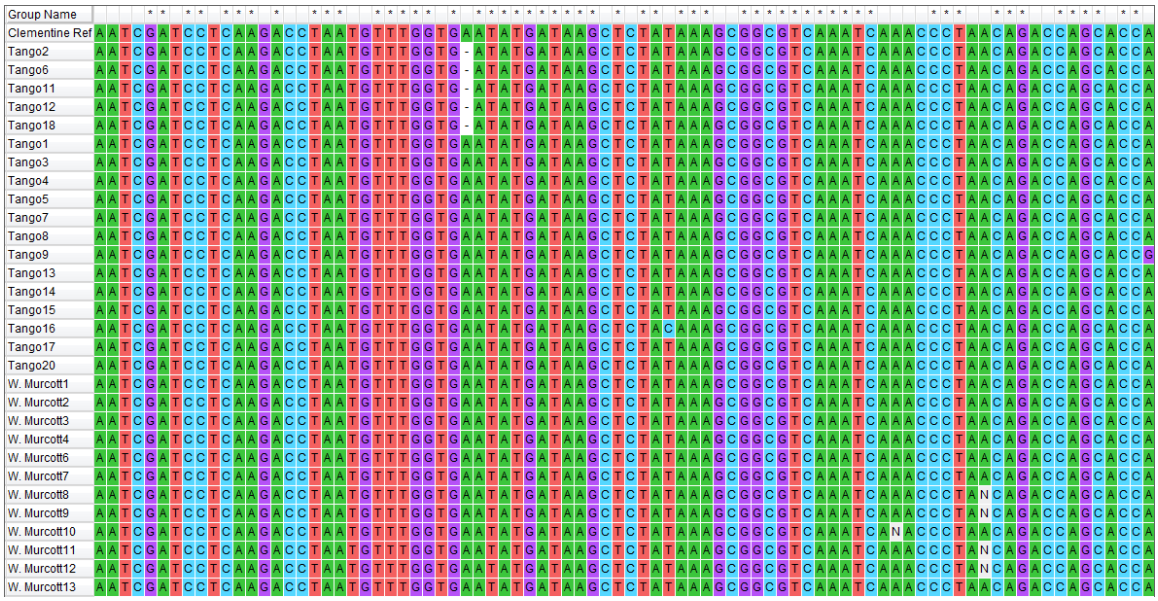


Figure 2.30 Alignment of Partial Sequences Including Del3-3 Indel from Sequencing Result of TA-Cloning. 18 sequences of Tango clones (sequence rows named "Tango" in the first column) and 12 sequences of W. Murcott clones (sequence rows named "W. Murcott" in the first column) aligned to *Citrus clementina* reference genome (first sequence row named "Clementine Ref").

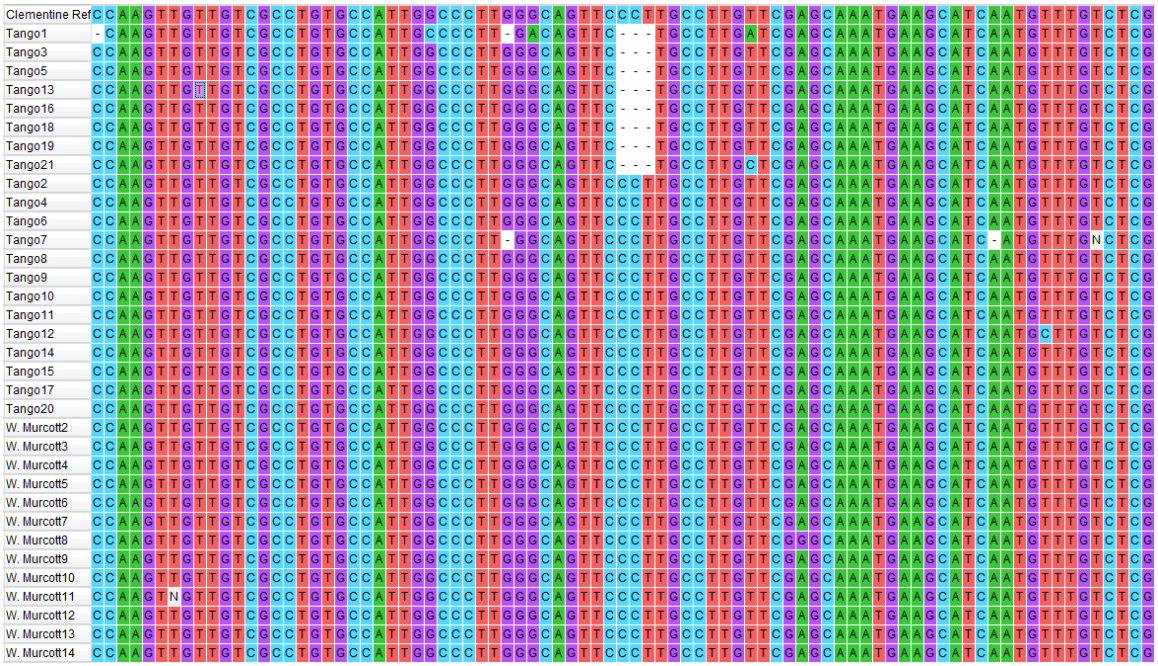


Figure 2.31 Alignment of Partial Sequences Including Del2-3 Indel from Sequencing Result of TA-Cloning. 21 sequences of Tango clones (sequence rows named "Tango" in the first column) and 13 sequences of W. Murcott clones (sequence rows named "W. Murcott" in the first column) aligned to *Citrus clementina* reference genome (first sequence row named "Clementine Ref").

Chapter 3: Chromosome Rearrangement of Tango

Abstract

Previous studies of meiotic chromosome spreads of W. Murcott and Tango showed that Tango chromosomes were misaligned in metaphase and segregated abnormally in anaphase, with quadrivalents and lagging chromosomes being observed, while W. Murcott meiosis was normal. It is likely that the chromosome rearrangements of Tango mandarin were induced during mutation breeding and cause the abnormal meiosis, the resulting gametes of which carry duplications and deficiencies that lead to pollen sterility and embryo abortion. The purpose of this study was to compare the Illumina NGS data and *Axiom® Citrus 15AX* SNP array data of W. Murcott and Tango leaf DNA to identify molecular signatures of chromosome rearrangements in the Tango genome. Comparative analysis of Illumina NGS data with novoBreak predicted candidate chromosome rearrangements. Copy number variation was analyzed by comparing the depth of coverage between Tango and W. Murcott along each chromosome. *Axiom® Citrus 15AX* SNP array data was used to analyze the loss of heterozygosity (LOH), which could reflect the existence of large heterozygous deletions, a special case of copy number variation. PCR with primers flanking the putative breakpoints and Sanger sequencing were used to verify the chromosome rearrangements and determine the genome position of chromosome rearrangement breakpoints. The depth of coverage and LOH analysis results suggested

there were no large chromosome segments duplicated or deleted in Tango. Analysis of rearrangements showed one heterozygous translocation of a 6 Mb segment on chromosome 2 and one heterozygous inversion of 4 Mb on chromosome 4 in Tango. The proximal and distal breakpoints of the apparent translocation on chromosome 2 were joined and the 6 Mb interstitial sequence may be translocated to a yet to be discovered genome location. In addition, a potential translocation joined the 31.5 Mb position of chromosome 3 and the 27.5 Mb position of chromosome 1. These three chromosome rearrangements could be responsible for the previously observed misalignment and abnormal segregation during meiosis, and the inviability of both male and female gametes in Tango, and contribute to its inability to fertilize other citrus varieties and the low seed content of Tango fruits.

Introduction

Structural variation of chromosomes includes abnormalities in chromosome structure, such as chromosome rearrangements resulting from large (>1kb) insertions, deletions, inversions, and copy number variations (CNVs) of DNA (KIDD et al., 2008; WECKSELBLATT AND RUDD, 2015). Chromosome rearrangements are biologically significant in certain instances such as human cancers and phase variation in *Salmonella* (LEE AND PHILLIPS, 1988; ROWLEY, 1982). There are different types of chromosome rearrangement, including chromosome fusion and/or fission, large duplications or deletions, non-homologous translocation and inversion of a chromosomal segment (RIESEBERG, 2001). If such genetic variations arise during meiosis because of recombination, they may be transmitted to the next generation (GAETA AND PIRES, 2010). The study of structural variations in human can help identify genes that are related to disease, breakage hotspots, parent-of-origin biases, as well as common mutational mechanisms (WECKSELBLATT AND RUDD, 2015). In human genetics, structural variations cause various chronic diseases and contribute to genetic diversity and evolution, and therefore have received much more attention than in plants, where there have been few thorough investigations of their role and consequences (SAXENA et al., 2014). The availability of genomes of many plant species in recent years and the reduced cost of genome sequencing and re-sequencing allow high-throughput genome-wide analysis for plant structural variations. Structural variations have been identified in several plant

species, some of which have been found to associate with phenotypic variation (SAXENA et al., 2014).

The formation of structural variations can result from two main mechanisms, non-homologous end-joining (NHEJ) and non-allelic homologous recombination (NAHR). NHEJ forms as the result of aberrant repair of uneven DNA double-stranded breaks and requires a relatively low sequence similarity at the breakpoints (SAXENA et al., 2014). NAHR relies on repetitive sequences in the genome and requires a high level of sequence similarity at the breakpoints (SAXENA et al., 2014). Repetitive sequences are prevalent in plants genomes, ranging from 10% (e.g., Arabidopsis) to more than 80% (e.g., bread wheat *Triticum aestivum*). Structural variations can also arise from duplication events or transposons, which also have a significant influence on genome evolution (SAXENA et al., 2014).

Structural variations can also result from either spontaneous or induced mutation. Irradiation can lead to large-scale genetic mutation in mammalian cells, such as structure alteration with a high proportion of gene deletions or rearrangements (KRONENBERG AND LITTLE, 1989; THACKER, 1986). Fast-neutron or X-ray irradiation can also cause inversions, insertions, deletions or more complex rearrangements ranging from 1 bp to more than 10 kb (GORBUNOVA AND LEVY, 1999). In plant breeding, irradiation such as gamma or fast-neutron is widely used for mutagenesis. Irradiation can inhibit seedling growth, reduce plant fertility, and induce chromosome aberration and micronuclei in root tip cells and pollen mother cells, with dose-dependent effectiveness (MEI et al., 1994). In rice root tip cells, the frequencies of micronuclei or chromosome aberrations were linearly

related to radiation exposure (MEI et al., 1994). If pollen mother cells were exposed to radiation during meiosis, the fertility of resulting seeds was closely related to ion damage (MEI et al., 1994).

Chromosome rearrangement involves chromosome breakage and rejoining. Double-strand breaks (DSB), which are the main lesion in genomes, can be induced by transposons, irradiation, endonucleases or mechanical reasons (GORBUNOVA AND LEVY, 1999). Double-strand DNA breaks initiate recombination and DSB repair process, which includes homology search, DNA synthesis and DNA repair, performed commonly via homologous recombination during meiosis (GAETA AND PIRES, 2010). There are two models which explain DSB repair, crossover and non-crossover: 1. double-strand break-repair (DSBR) model, which explains the mechanism of reciprocal exchange and gene conversion during meiotic recombination (SZOSTAK et al., 1983); and 2. synthesis-dependent strand annealing (SDSA) model, which explains gene conversions (GAETA AND PIRES, 2010). In plants, DSBs are induced by *Spo11* homologs during leptotene of meiosis. Single-strand DNAs exposed due to DSBs act as homology search targets (GAETA AND PIRES, 2010). Recombination can also occur between repeats within the same chromosome, resulting in intrachromosomal exchange, or repeats on homologous chromosomes, or non-homologous chromosomes that share some homology (GAETA AND PIRES, 2010). During meiosis of an individual with a heterozygous rearrangement, recombination between chromosomes carrying different rearrangements often generates genetically unbalanced gametes (containing duplications or deficiencies), which could lead to gamete or zygote inviable.

The result brings about the selective recovery of non-recombinant chromosomes in viable offspring and an effective reduction in recombinant chromosomes (RIESEBERG, 2001).

Efficient double-strand break repair is of importance for the survival of all organisms. Double-strand breaks can be repaired via the homologous recombination (HR) pathway and the non-homologous-end joining (NHEJ) pathway (ROY, 2014). In bacterial and yeast, HR is commonly utilized, in which the non-damaged copy of the homologous DNA duplex is required to serve as a template to form the heteroduplex for repairing the damaged strand (ROY, 2014). In plants, HR is essential for DSB repair during the early stages of gamete formation where homologous chromosome pairing and recombination are induced by DSB, while the NHEJ pathway plays an essential role in a majority of DSB repair, where the broken ends of double-stranded DNA are joined directly regardless of sequence homology (ROY, 2014). The reasons for the DSB repair being error-prone include when both DNA strands are damaged and no complementary strand is available to recover missing information, when both DNA strand ends are not compatible for ligation, or when breaks are enlarged by exonuclease activity (GORBUNOVA AND LEVY, 1999). Thus, the rejoining of broken double-strand DNA might lead to loss of genetic information and various genomic rearrangements (GORBUNOVA AND LEVY, 1999; PIPIRAS et al., 1998). During DSB breakage-induced replication, a damaged single strand from a replication fork invades sister or non-sister chromatids and uses them as a template to repair DNA (SUNG AND KLEIN, 2006). Then gene conversion is generated due to nonreciprocal genetic material transfer. The errors produced by DSB repair in NHEJ pathway can lead to

insertions and deletions at the site of ligation, or a translocation could be generated by ligating unlinked DNA (GAETA AND PIRES, 2010; PIPIRAS et al., 1998; PUCHTA, 2005).

However, the DSBs and structural variations are not arbitrary in the chromosomes. Nonrandom distribution of chromosome breakpoint locations was found for both spontaneous and induced breakage. It was reported that chromosomes containing the nucleolus organizer region (NOR) have a high probability to be involved in structural alterations (LEE AND PHILLIPS, 1988). Previous analysis of position and distribution of chromosome rearrangement breakpoints in cultured plant cells showed that some specific chromosome regions, especially regions containing heterochromatin, are frequently affected by structural variation (ROWLEY, 1982). In tomato (*Solanum lycopersicum*), reciprocal translocation breakpoints and deletions were found preferentially within the heterochromatin region of chromosomes (GILL et al., 1980; KHUSH AND RICK, 1968). In maize (*Zea mays*), reciprocal translocation breakpoints and inversions were found preferentially near the centromere, and translocation breakpoints were located in heterochromatic regions more frequently than euchromatic regions (JANCEY AND WALDEN, 1972; LONGLEY, 1961). In *Vicia faba*, there was a high chance of breakage within heterochromatic regions (RIEGER et al., 1975).

At the genetic level, structural variation can lead to the disruption or fusion of genes and thus abnormal development if the breakpoints are located in open reading frames of genes (WECKSELBLATT AND RUDD, 2015). Fusion genes can result from deletions, duplications, translocations, and chromothriptic rearrangements (WECKSELBLATT AND RUDD, 2015). Disrupted or fused genes at the breakpoints are good candidates for human disease or new

gene functions that are related to new phenotypes because the rest of genome is intact (WECKSELBLATT AND RUDD, 2015). The structural variations can also cause position effects for the intact genes that are near the breakpoints, altering gene expression (WECKSELBLATT AND RUDD, 2015).

At the genome level, the consequences of chromosome rearrangement could include increases in chiasma frequency. In *Hypochoeris radicata* two rearrangements, a chromosome fission and a balanced chromosome rearrangement, showed a direct correlation between the increase of chiasma frequency and the increase in the number of potentially-paired segments generated by the rearrangement at pachytene (PARKER, 1987). Individuals with a heterozygous rearrangement formed more chiasmata than structurally-normal individuals in families of this species. Also, chromosome rearrangements resulting from error-prone DNA repair may also provide the raw material for plant evolution, and the genomic changes can be transferred to the next generation (GORBUNOVA AND LEVY, 1999). Some studies of the selfed progeny of different viable translocation heterozygotes of *Nigella damascena* L. caused by gamma irradiation showed a ring or a chain of four chromosomes configuration in some meiocytes, and the pollen fertility exhibited a negative but non-significant correlation with seed number per plant (ANINDITA AND DATTA, 2006). Defective female gametogenesis resulted in the total failure of seed setting in heterozygotes (ANINDITA AND DATTA, 2006). Also, the formation of a ring was found when the interchanged chromosome parts were long and the breakpoints were closer to the centromere, while the formation of the chain was found when the interchanged chromosome parts were relatively short (ANINDITA AND DATTA, 2006).

In order to detect chromosomal variation, a variety of techniques have been employed. Chromosome banding analyzed by microscopy is the traditional method that detects large deletions, duplications, inversions and translocations (WECKSELBLATT AND RUDD, 2015). It is more effective in species with relatively large chromosomes than in those with small chromosomes where light microscopy lacks sufficient resolution. In the 1980s, the most penetrating analyses used light microscopy to observe stained squash preparations of cells (LEE AND PHILLIPS, 1988). Combined with simple staining techniques, chromosome banding techniques not only permitted observations of chromosome number and other cytological features such as chromosome length, centromere position and arm length ratio, but also revealed the linear differentiation of chromosomes and some more sophisticated analysis such as the distribution of heterochromatic regions (LEE AND PHILLIPS, 1988). With regards to meiotic analysis, it provides advantages for evaluating homologous chromosome pairing behavior through different stages of meiosis (LEE AND PHILLIPS, 1988). Chromosome fusion events could be observed as chromosomes exhibiting two or more centromere-like constrictions (LEE AND PHILLIPS, 1988). However, chromosome banding is incapable of detecting small microdeletions and microduplications (WECKSELBLATT AND RUDD, 2015). Other techniques which can detect at least some types of chromosomal variants include fluorescence *in situ* hybridization (FISH), which can verify the location of chromosomal segments or the position of specific DNA sequence probes, whole-genome sequencing (WGS), and array comparative genome hybridization (CGH), which detects changes in copy-number and allelic ratios such as deletions and

duplications but cannot determine the location and organization of structural variations (WECKSELBLATT AND RUDD, 2015).

The recent development of next-generation sequencing (NGS) technology has been assisting the detection of copy-neutral and copy-number-changed structural variation by utilizing sequence read depth and paired-end reads that span the breakpoints (WECKSELBLATT AND RUDD, 2015). Chromosome banding and FISH can also assist in verifying complex structural variations identified by NGS (WECKSELBLATT AND RUDD, 2015).

NGS can be applied to structural variation studies through whole-genome, whole-exome and whole-transcriptome approaches (MEYERSON et al., 2010). By using NGS methods, the efficiency and resolution for detecting different types of somatic genome alterations, including nucleotide substitutions, small insertions and deletions, copy number variations, chromosome rearrangements and microbial infections, have been significantly increased (MEYERSON et al., 2010). Detection of breakpoints of large-scale chromosome rearrangements at the sequence level, as well as analysis of genes that are disrupted, fused or misregulated, become possible owing to NGS (WECKSELBLATT AND RUDD, 2015). The mutational mechanisms of chromosome rearrangement can be revealed by analyzing the DNA sequence at structural variation breakpoints, where NHEJ repair is represented as absence of sequence homology while NAHR repair is represented by long stretches of homologous sequence shared between breakpoints, and DNA replication-based mechanisms are reflected as inserted or inverted sequences at breakpoints (WECKSELBLATT AND RUDD, 2015). The NGS technologies not only make it easy to detect some genomic

alterations with a low frequency in samples, but also make it feasible to discover novel chromosomal alterations at a high resolution, as shown in human cancer studies (MEYERSON et al., 2010). In addition, by using NGS, multi-fold coverage and high-quality data are possible. The over-sampling of the genome provides enough signal coverage to overcome experimental noise and allow detection of genome alterations in heterogeneous samples (MEYERSON et al., 2010).

WGS provides the most comprehensive genome characterization and is ideal for analysis of copy-neutral segments with few breakpoints within repeats, such as chromothripsis in human disease, which is many chromosome rearrangement breakpoints clustered on a single chromosome (WECKSELBLATT AND RUDD, 2015). Complex structural variations such as tens of breakpoint junctions across multiple chromosomes are difficult for short-read WGS to deal with (WECKSELBLATT AND RUDD, 2015). For library construction, DNA is firstly fragmented into 200-400 bp and next-generation sequencing yields reads from both ends of a fragment, making paired-end reads (MEYERSON et al., 2010). Reads with the expected distance between paired ends are placed on the reference genome and discordantly mapped reads are used for detection of structural anomalies (MEYERSON et al., 2010). Another technique, mate-pair sequencing, provides sequences of ends from larger DNA fragments that are 2-5 kb in length, increasing the capability and accuracy of detecting large structural variations, especially for detecting breakpoints contained in interspersed repeats which are difficult to identify by standard NGS methods, and inversions and balanced translocations which cannot be detected by copy-number methods (WECKSELBLATT AND RUDD, 2015). However, the short-read approaches are restricted by

highly complex genomes which include many long repetitive elements, copy number alterations and structural variations, due to the insufficient read length (GOODWIN et al., 2016). Newly developed sequencing technologies support long-read sequencing with length up to 100 kb, permitting the analysis of large complex structural variations and large repetitive regions (GOODWIN et al., 2016). There are two major types of long-read sequencing technologies: 1. Single-molecule real-time (SMRT) technology, applied by Pacific Biosciences and Oxford Nanopore platforms, which offers sequencing of reads with over 10 kb in length that can cross repetitive elements and is suitable for GC-rich regions and tandem repeats, at the cost of higher error rate per read. 2. Synthetic approaches, which relies on existing short-read technologies to construct long reads *in silico* (GOODWIN et al., 2016). By using those NGS approaches, for example, the copy number variant can be estimated by counting the number of reads at a genomic locus of both mutant and normal samples and the mutant-to-normal ratio is calculated. Mutant genomic regions with high coverage or low coverage compared to normal are candidates for copy number alterations (MEYERSON et al., 2010). For chromosome rearrangements, multiple types of rearrangements can be detected based on insert size and read orientation, including intrachromosome rearrangements (e.g., inversions, tandem duplications, large insertions/deletions, etc.), reciprocal and non-reciprocal interchromosome rearrangements, and more complicated rearrangements such as combinations of these events (MEYERSON et al., 2010).

W. Murcott mandarin, also known as Nadorcott or Afourer, is originally from Morocco and was imported to California in 1985 (ROOSE AND WILLIAMS, 2007). It is believed that

Murcott tangor is an F1 hybrid of a cross between an unknown mandarin and a sweet orange, and W. Murcott is a chance zygotic seedling from the Murcott tangor, having an unknown pollen parent. This grandparent-grandchild relationship between sweet orange and W. Murcott was supported by sequence analysis, which demonstrated that 34% of the sweet orange genome is shared with W. Murcott (WU et al., 2014). W. Murcott is widely planted throughout California due to the high quality of its fruits, which have deep orange color and sweet juice, and are easily peelable (ROOSE AND WILLIAMS, 2007). In addition, a significant feature is they can be seedless due to self-incompatibility, in which the male gametes are unable to fertilize female gametes, preventing seed formation and resulting in facultative parthenocarpy (GAMBETTA et al., 2013; OLLITRAULT et al., 2007). In order to set seedless fruits of W. Murcott mandarin, either the trees need to be planted in isolated areas or whole rows must be covered with screen during flowering season to prevent cross-pollination, otherwise they will produce seedy fruits as a result of cross-pollination by bees-transferring pollen from other Citrus with high pollen viability, which significantly reduces the fruit value (CROWLEY, 2011).

Tango mandarin is a low-seeded variety derived from gamma irradiation of budwood of diploid mandarin cultivar W. Murcott (Afourer) mandarin, which may be the same as Nadorcott of Europe. Budwood of W. Murcott from Lindcove, California was treated with 50 Gray units of ^{60}Co gamma at Riverside, California and the irradiated buds were propagated on various rootstocks in a greenhouse and then planted in the field for fruit production and evaluation (ROOSE AND WILLIAMS, 2007). One tree propagated on C32 citrange rootstock showed distinctive characteristics including fruit with very low seed

counts in all situations of cross-pollination and the same high quality and production as on the original 'W. Murcott' cultivar (ROOSE AND WILLIAMS, 2007). This tree then selected for further trials. After multiple rounds of propagation and testing, this selection was named 'Tango' and commercially released.

It is hypothesized that irradiation of W. Murcott budwood caused DNA double-strand breaks (DSBs) followed by DSB repair. Since DSB repair is error-prone, misrepair can lead to chromosomal structure changes such as chromosome rearrangement and large deletions through various DSB repair pathways (SANKARANARAYANAN et al., 2013). In addition, homologous chromosome misalignment, chromosome segregation failure and lagging chromosomes during meiosis as observed in previous studies also indicated the potential for chromosome organization being changed by gamma irradiation (CROWLEY, 2011). Large deletions and other chromosome rearrangements could be used as markers to differentiate Tango from W. Murcott. The objective of this study is to identify chromosome structural variations in Tango compared to its W. Murcott parent, and better explain the meiotic abnormalities observed in Crowley's studies. Also, structural variations could enable development of fast, robust, and low-cost PCR-based methods for discrimination of Tango from W. Murcott in addition to the existing phenotype markers.

In order to analyze chromosomal abnormalities, powerful computational tools are required to deal with the huge volume of data produced by next-generation sequencing. For normal-mutant pair analysis, the following are usually required: analyze data simultaneously from both mutant and normal to identify rare events such as SNP frequency, analyze very different and highly rearranged genomes, and analyze mutant samples with unknown levels

of contamination from normal samples (MEYERSON et al., 2010). Firstly, the next generation sequencing data must be aligned to the specific chromosome position and strand of a relevant reference genome before analysis. Tools such as Bowtie, SOAP2 (Short Oligonucleotide Analysis Package) and BWA (Burrows-Wheeler Aligner) differ in accuracy of mapping noisy reads, long and short reads and computational efficiency (MEYERSON et al., 2010). Secondly, mutation detection analysis should utilize tools with low false-positive rates since genome alterations are rare. The allelic fraction, which is the expected fraction of mutant reads among all reads in the same genomic position, can be affected by contamination, experimental or alignment bias, ploidy and copy number (MEYERSON et al., 2010). Thus the false-positive and false-negative rates depend not only on coverage but also the allelic fraction (MEYERSON et al., 2010). Thirdly, the detected mutations should be validated by experimental approaches, the accuracy of which may be affected by high frequencies of false-positives and false-negatives. For SNPs, TaqMan® assay is a common genotyping method while Pyrosequencing and array-based technologies are suitable for large-scale SNP analysis with high efficiency and accuracy (GUNDERSON et al., 2005; HUI et al., 2008; NORDFORS et al., 2002). Indels can also be assayed by TaqMan® method or simply PAGE-based genotyping method by separating the homoduplex and heteroduplex DNA (BARKLEY et al., 2010; ZHU et al., 2014). For rearrangements, PCR amplification of the region surrounding the breakpoints followed by sequencing is required, but this method is not high-throughput (MEYERSON et al., 2010).

NovoBreak is an algorithm that produces local assemblies for the discovery of structural variation breakpoints from NGS data. It outcompetes other existing algorithms on real

cancer genome data in both performance and accuracy (CHONG et al., 2017). It is also highly sensitive in identifying short insertions and deletions by detecting discordant reads spanning breakpoints (CHONG et al., 2017). Instead of firstly aligning paired-end reads to a reference genome accurately or relying on local assembly of aligned and partially aligned reads of previously identified candidate structural variations, novoBreak omits perfectly aligned reads and clusters imperfectly aligned reads according to the similarity of a set of short nucleotide stretches present in the subject genome rather than the reference genome, which does not require the high accuracy of read alignment, then identifies breakpoints by comparing the resulting high-quality contigs that are assembled from clusters of read pairs spanning breakpoints to the reference genome sequence (CHONG et al., 2017). Also compared to whole-genome assembly, novoBreak is neither computationally intensive nor affected by repeats, read length, and coverage (CHONG et al., 2017). Since the *Citrus clementina* reference genome assembly is incomplete, and differences exist between W. Murcott/Tango and *Citrus clementina* and thus read alignment may be inaccurate, novoBreak is more suitable than other algorithms for our studies.

Materials and Methods

Paired-End Sequencing and Read Alignment

Illumina paired-end libraries were prepared from genomic DNA of Tango (Field 1B-R28-T02) and W. Murcott (Field 12B-R48-T03) mandarin trees at UCR. DNA was fragmented mechanically by sonication to fragment lengths of ~ 300 bp. Size selected DNA fragments were purified from e-gels and checked for fragment length distribution on Agilent DNA 7500 chips. Paired-end library preparations were performed according to the manufacturer's instructions (Illumina "Preparing Samples for Paired-End Sequencing", Part # 1005063 Rev. A June 2008). One lane of each library was sequenced on an Illumina Genome Analyzer II at the Institute for Integrative Genome Biology at the University of California, Riverside.

Sequence alignment was done by Prognosys Biosciences (San Diego, CA) using BWA and proprietary algorithms. Both W. Murcott and Tango Illumina data was aligned to the *Citrus clementina* haploid genome assembly v1.0. The aligned data was then compressed to a binary format (bam file), sorted and indexed by using samtools sort (http://samtools.sourceforge.net) with default settings.

Data Filtering

Due to the nature of library construction using amplification during next-generation sequencing, duplicated reads are inevitable. In order to minimize the effect on coverage of specific positions, duplicated reads were marked from the raw data by using Picard MarkDuplicates (https://broadinstitute.github.io/picard) with default settings, which compares five prime positions of both reads and read pairs, followed by removal of marked duplicate reads by using the REMOVE_DUPLICATES command

```
java -Xmx32G -jar picard.jar MarkDuplicates
INPUT=input_sorted.bam
OUTPUT=output_sorted_PicardMarkDuplicates.bam
METRICS_FILE=metrics.txt REMOVE_DUPLICATES=true
```

Copy Number Variations

Two approaches were used to detect copy number variation between W. Murcott and Tango. The first used the genome sequence data. Since both W. Murcott and Tango reads were mapped to the Clementine reference genome, we divided the reference genome into non-overlapping windows with different sizes (0.1kb, 1kb and 10kb) for different resolution analysis. For each window, average coverage was calculated as the sum of each base's coverage divided by the window size. Window coverage of both W. Murcott and Tango data was plotted along each chromosome. In order to reduce the effect of repeats in both W. Murcott and Tango genomes that are not correctly assembled in the reference

genome, the log2 ratio of average coverage of Tango versus W. Murcott for each window of each resolution was calculated and plotted along each chromosome.

The second method utilized hybridization to the *Axiom® Citrus 15AX* SNP array (Hiraoka et al., in prep.) to conduct loss of heterozygosity (LOH) analysis. Single samples of W. Murcott and Tango DNA were hybridized to the arrays by Affymetrix. Probes classified as polymorphic-high-resolution or monomorphic-high-resolution by Axiom® Analysis Suite were analyzed. For W. Murcott and Tango data, positions called as either heterozygous or homozygous for both W. Murcott and Tango, and positions called as heterozygous for W. Murcott but homozygous for Tango, which are LOH positions in Tango, were collected. For each window of 0.1 and 1 Mb size along the chromosome, the ratio of LOH position counts in Tango within the window versus heterozygous position counts in W. Murcott within the window was plotted along the chromosome. The region (more than one continuous window) where there is a high Tango LOH/W. Murcott heterozygosity ratio with a sufficient density of LOH positions, combined with reduced depth of coverage of Tango NGS data, may indicate the region is lost in one of the two haplotypes in Tango compared to W. Murcott. Markers in each comparison class (homozygous position for both W. Murcott and Tango, the heterozygous position for both W. Murcott and Tango, LOH position in Tango compared to W. Murcott) were plotted separately along each chromosome.

Chromosome Rearrangement Prediction

For chromosome rearrangement, novoBreak was used to analyze the data. Both W. Murcott NGS data and Tango NGS data were used as input for chromosome rearrangement analysis, and *Citrus clementina* v1.0 genome assembly was used as the reference genome. To run novoBreak, the following parameters were used:

```
./run_novoBreak.sh novoBreak_distribution_v1.1.3rc
Clem1.0.fa Tango_PicardMarkDuplicates_sorted.bam
WMurcott_sorted_PicardMarkDuplicates_sorted.bam 8
```

Only predicted deletions and inversions that were large enough to cause abnormal chromosome pairing during meiosis (>1 Mb) were kept for the following analysis. Then the predicted breakpoints of each retained predicted large deletion, inversion and translocation were investigated using the read alignment display in IGV.

For each predicted chromosome rearrangement, if abnormal read pairs were found at the breakpoints and the chromosomes involved matched the predicted rearranged chromosomes, the chromosome rearrangement was retained for PCR verification. Several classes of abnormal read pairs were retained as follows. For interchromosome rearrangement: 1: the two ends of a pair were mapped to two different chromosomes, or 2: the full length of one end and at least 16 bases of the proximal part of other end were mapped to positions within the insert size on one chromosome, while the distal part was aligned to a different chromosome with at least 16 bases. For intrachromosome rearrangement: the two ends of a pair were mostly mapped to two positions separated by

at least 1 Mb on the same chromosome of the reference, which significantly exceeds the range of the average insert size, or one end was mapped with full length to a position and the proximal part of the other end was aligned to a nearby position within the insert size while the rest was aligned to a position that exceeds 1 Mb distance on the same chromosome. The criteria for differentiating between a large deletion and a large inversion depend on the alignment orientation of a group of such abnormal read pairs.

Primer Design for Real-Time PCR Verification

For each chromosome rearrangement, the sequences of chimeric paired-end reads which themselves or for which the sequence between two ends included the breakpoint were firstly collected, as possibly representing actual sequences in the genome. To obtain a larger sequence range for primer design, based on sequence overlap and relative mapping coordinates of reads, the sequences were manually assembled into a pair of longer sequences, which represented two ends of a read pair assembly. The breakpoint was therefore included in the assembled sequence within either end or in the unknown sequence between two ends. Then the assembled sequences of the two ends were BLAST searched against the *Citrus clementina* reference genome to verify the predicted breakpoint. Forward and reverse primers were then designed according to the assembled read sequences flanking the breakpoint by using web-based Primer3 (http://bioinfo.ut.ee/primer3-0.4.0) and PrimerQuest Tool (https://www.idtdna.com/primerquest). Multiple primer pairs were

designed based on the number of assemblies (Table 3.1), and an MDH gene primer pair Cit1249 + Cit1250 (unpublished sequence) was used as positive control.

Real-Time PCR Validation

Real-time PCR was used to validate predicted rearrangements based on detection of sequences predicted as unique to Tango. Real-time PCR reactions included: 1X SsoAdvanced™ Universal SYBR® Green Supermix (Bio-Rad, cat no. 172-5270), 10 µM forward primer, 10 µM reverse primer, 2 µl W. Murcott or Tango genomic DNA of various concentrations and ddH_2O to 20 µl total reaction volume. Reactions were placed in BIO-RAD Multiplate™ 96-Well PCR Plates (cat no. #MLL9601), and plates were covered with USA Scientific X-Pierce cross-cut pierceable film (cat no. 2997-0100) during preparation to prevent cross-contamination. After preparation, the plate was sealed with BIO-RAD Microseal® 'B' PCR Plate Sealing Film (cat no. #MSB1001) and Film Sealing Roller for PCR Plates (cat no. #MSR0001), and placed in BIO-RAD CFX96 Touch™ Real-Time PCR Detection System programmed with the following settings: amplification stage: initial 95 ℃ step for 3 min; 95 ℃ denaturation step 10 sec; 58 ℃ annealing step 30 sec; repeat denaturation and annealing steps 45 times, followed by melt curve stage from 65 ℃ to 95 ℃ with 0.5 ℃ increment at 5 sec/step rate. Data was collected by BIO-RAD CFX Manager™ Software (Version 3.1).

Sanger Sequencing across Putative Breakpoints

Four pairs of primers for conventional PCR were designed for each deletion and inversion breakpoint. Each pair of primers were designed with an expected amplicon length about 400-600 bp (Table 3.2). PCR products were then isolated and direct Sanger sequenced at Institute for Integrative Genome Biology (IIGB) instrumentation facilities of University of California, Riverside with both corresponding forward and reverse primers (Table 3.2) as sequencing primers.

For large deletion DEL2, there were two local read assemblies of the breakpoint of scaffold_2, namely, local assembly 1 and local assembly 2. Local assembly 1 represented the sequence that was separated by the deleted part with respect to the reference genome. Deletion 2-1 primer set was thus designed flanking the large deletion (nBDEL_LG_F and nBDEL_LG_R in Figure 3.14 and Table 3.2). The BLAST result of the sequence of local assembly 2 showed that it was the join of sequences within the "deleted" region flanked by the two breakpoints. Deletion 2-2 primer set was thus designed to flank the same breakpoint with the sequence inside the deleted region (nBDEL_SM_F and nBDEL_SM_R in Figure 3.14 and Table 3.2).

For large inversion INV4, two primer sets were designed flanking the two breakpoints of the inversion respectively on scaffold_4 (nBINV_LFT_F and nBINV_LFT_R for Inversion 4-1, and nBINV_RGT_F and nBINV_RGT_R for Inversion 4-2, respectively, in Figure 3.15 and Table 3.2).

Results

No Copy Number Variation of Large Chromosome Segment Found

The results of the log2 ratio of Tango coverage versus W. Murcott coverage showed that on all nine chromosomes this value was between -0.30 to 0.35. No island (i.e., continuous high ratio) or valley (i.e., continuous low ratio) was observed on the plot with a ratio close to or larger than 1 or less than -1, indicating the coverage between Tango and W. Murcott along all chromosomes was not largely different (Figure 3.1). Although some extreme values deviating from the average were still observed in the form of thin peaks or valleys on the plot, none of them are larger than 100 kb, which is unlikely to cause sterility.

W. Murcott and Tango data from the *Axiom*® *Citrus 15AX* SNP array were also analyzed. Only high-resolution calls of nuclear genome probes present in both W. Murcott or Tango were retained for data analysis. Only 727,224 nuclear genome probes classified as polymorphic high-resolution and 137,147 as monomorphic high-resolution were analyzed. For each Mb of the chromosome, the total number of markers that are heterozygous in W. Murcott and the number of positions that are heterozygous in W. Murcott but homozygous in Tango (i.e., LOH positions) were counted. LOH counts per Mb each chromosome (LOH density), number of markers per Mb that are heterozygous in W. Murcott (heterozygous positions) and the ratio of those two counts in percentage form (LOH %) were plotted along each chromosome (Figure 3.2 and Figure 3.3). The plots showed maximum LOH per Mb on all chromosomes was 7, compared to 290 LOH within a Mb in Valencia (in an area

confirmed as containing a heterozygous deletion) within the same SNP array analysis. Also, no successive 1 Mb windows with high LOH percentage were observed. The maximum percentage of LOH was between 19 Mb – 20 Mb of chromosome 2, which is only one LOH out of two heterozygous markers within 1 Mb, only very weak evidence of a deletion because of the small number of heterozygous markers in this interval. The next most frequent percentages of LOH were at 35 Mb of chromosome 5, which had about 44% LOH SNPs, but W. Murcott had only nine heterozygous SNPs in this interval. At 20 Mb – 22 Mb of chromosome 6, there is about 21% to 37.5% LOH in intervals with 13 to 19 heterozygous markers in W. Murcott. At 23 Mb of chromosome 2, there is a 25% LOH while W. Murcott had only 12 heterozygous markers in this interval. The percentage of LOH in other all regions and other chromosomes was lower than 10%.

The density of valid markers on this SNP array, the nuclear genome probes exhibiting heterozygous state in W. Murcott and the LOH density of Tango in each 100 kb window were plotted along each of the nine chromosomes (Figure 3.4). No large (> 1 Mb) region containing many markers heterozygous in W. Murcott but homozygous (or hemizygous) in Tango was observed.

Three Chromosome Rearrangement Discovered

After removal of duplicate reads by Picard, W. Murcott and Tango NGS data was analyzed with novoBreak to predict potential chromosomal abnormities. Totally 29 structural variation events with respect to *Citrus clementina* reference genome were predicted,

including 12 large deletions, five translocations and 12 large inversions. Since the chromosome rearrangements have to be sufficiently large to be causative, two large deletions, all five translocations and three large inversions were retained (Table 3.3). After inspection of the aligned read sequences at each breakpoint with IGV, three predicted chromosome rearrangements were retained for PCR validation: large deletion DEL2 (namely DEL2-6572073:12563420), translocation TRA1/3 (namely TRA1/3-27525574:31452544) and large inversion INV4 (namely INV4-3540415:7515397) (Table 3.4). The other four predicted translocations were excluded from qPCR validation because no primer sets that had single hits in the genome could be designed or the PCR had multiple products, and large inversion INV4 and large inversion INV4-3540417:7515406 were considered as the same inversion event because their breakpoint coordinates were similar. Only INV4 was retained.

For large deletion DEL2, there were two local read assemblies obtained from chimeric read pairs due to effects of a chromosome rearrangement (called a deletion by novoBreak). Local assembly 1 consisted of two groups of assembled sequences in which some reads and read-pairs spanned the apparent deletion breakpoints at 6572073 bp and 12563420 bp on scaffold_2 of the reference genome. One group of reads mapped to the positive strand and the other to the negative strand of the same DNA segment. Primer pair NB81 and NB82 flanking the potential breakpoint were thus designed to amplify a segment corresponding to the group with chimeric end mapped to the positive strand. Similarly, primer pair NB87 and NB88 were designed to amplify the sequence with chimeric end mapped to the negative strand. Local assembly 2 was a DNA segment that did not overlap local assembly

1 and aligned between the two DEL2 breakpoints. Primer pair NB93 and NB94 were thus designed to amplify a segment corresponding to assembly 2 (Figure 3.12, Table 3.1).

DEL2 local assembly 1:

ATAACAAAAAAGAAATTTAATTGAGGAAAATCAAATATCTATAAAGATTTGTACCAATC

ATGTACACAAATCAAACTCCCTCAGTGTTCTAAAGTGTAAAG**G**CCAGAGAAGACCTTGA

CTCACTATTGTGTTTGAAAAC**G**AGTGTTGCTA**T**CAACGCGGGAAGAAGATGATTATGGG

CCTGCTCATGGTGTCCTTTTCCTGGGTTAAAGTGTCTATATTTTAAATTTTGAAGGGAT

TGATGGAGTAGGACTTGGACTGGTTAGGCACCATTACTCGCTTGGAAAGTAAGTAATGT

CCAAGATTTGAGTGCTTTATACCATCAATCATTGATAATCAGAACTGTTGAGAAGTTTT

CGAAATTGAAGCCATCTTGATGTGGACATCCAATGATTCACCATGTATG

in which the first 218 bases (shown in black) match the positive strand of scaffold_2 from 6571856th bp to 6572075th bp with mismatches of two G's and a single nucleotide T at positions where the reference has two A's and a dinucleotide TT respectively. The second portion of this assembly includes 188 bases (shown mostly in red) which match the positive strand of scaffold_2 from 12563421st bp to 12563608th bp, with the first three bases (TAT) of the latter part overlapping the last three bases of the former part.

The above local read assembly represented the sequence affected by DEL2 with two breakpoints at 6572075 and 12563421. There was no evidence of overlap with the other local assembly (DEL2 local assembly 2) so these two local assemblies could not be combined.

DEL2 local assembly 2:

Chimeric end of DEL2 local assembly 2:

ATTACAGCCCATACAAGGGCCGCCAATCTGAAAATGAAGGAAGTCGGATCCATGATTGT
AAATTTTCGGTTGATTGGATTCTGCAAGTTCTTATAAAGTTCATAAAAAAAATGAATAA
TTCATGCTATTTCCACTCAGAGCTTCAAAG

in which bases 8 to 103 (black) match the negative strand of scaffold_2 from 6572169th bp to 6572074th bp, and bases 112 to 149 (red) match the negative strand of scaffold_2 from 12563412th bp to 12563376th bp. The stretch of eight A's in between did not align due to the low complexity.

The non-chimeric end of DEL2 local assembly 3:

TTCATAAAAGGCATTGCACATATGAAATGGTAAAAGATTAATAGATTGGTTTATAATGC
TGTAGTATGTTATATATGTTTTGTCATTATAGCTCAATTGTTGATTTGCCAATTTAGAA
TTTATTTGTTGCTGGAGTATTTTCTTTTTTCAAATGGT

matches the positive strand of scaffold_2 from 12563210th bp to 12563365th bp.

For translocation TRA1/3, only one local read assembly was obtained from four chimeric read pairs that were affected by a translocation without reciprocal read pairs. The local assembly consists of two sequences, which represent read pairs with one end aligned to scaffold_1 and the other end aligned to scaffold_3, flanking the potential breakpoints on both chromosomes. Primer pair NB13 and NB14 were thus designed according to either end of the assembly to amplify a segment spanning the potential breakpoints (Table 3.1).

End of TRA1/3 local assembly aligned to scaffold_1:

TGATTAGCCTTAGCAATAGGCAGAGTACTCTATTTGAAAGTACCATTTCTATTACTATT
TGCTTTCGAATTTCTTTGAGAATCTCTCAGGAACTGGGCAAAGGTATGTTGAGACAAGC
ACCCCCAATG

the whole sequence of which matches the positive strand of scaffold_1 from 27525397th bp to 27525524th bp.

End of TRA1/3 local assembly aligned to scaffold_3 had many regions of low complexity:

AAATARAAATTAAAAAAAAAAATTTTTTTGTTTTTTTTGTTTTTATGTGAAAAGCCAAT
GATAAGAATTTGATTGACAATGAATAT**A**GTTTTTAGAAAGGGTTGAATATTGTCTTAGT
TTATACT

the best hit was

ATGTGAAAAGCCAATGATAAGAATTTGATTGACAATGAATAT**G**GTTTTTAGAAAGGGTT
GAATATTGTCTTAGTTTATACT

with a G instead of an A, which matched multiple positions in the *Citrus clementina* genome assembly, including the positive strand of scaffold_3 from 31452459th bp to 31452539th bp.

However, no reciprocal read group that crossed scaffold_3 and scaffold_1 around the predicted breakpoints was found.

For large inversion INV4, there were two local read assemblies obtained from chimeric read pairs that were affected by an inversion. Both local assemblies have one chimeric end including a breakpoint while the other normal end did not. Primer pair NB143 and NB144

were thus designed according to the sequences of two ends of local assembly 1 to amplify a segment spanning the potential breakpoints. Primer pair NB115 and NB116 were designed to amplify a segment spanning the potential breakpoints according to the sequence of one end of local assembly 2 that includes the breakpoint, while primer pair NB153 and NB154 were designed for local assembly 2 in a similar pattern to primer pair NB143 and NB144 (Figure 3.13, Table 3.1).

INV4 local assembly 1:

The chimeric end of INV4 local assembly 1:

TGGCACCGATCGACCCGCACTCGTTCACCGAGTCAACTCACCCGCTGACCACACACATC
TCCCTCTCTCTCTACTTCGACTTCTCTTCATCCACCATCCACGCCGCCGCCATCCTTAC
CCTCGCCTCCCCAATATTGGTGGTTATCGACAAC

in which the first 128 bases (black) plus adjacent CCA match the positive strand of scaffold_4 from 3540285th bp to 3540415th bp, and the last 24 bases (red) match the negative strand of scaffold_4 from 7515397th bp to 7515374th bp, where the first three bases (CCA) of the latter part overlaps the last three bases of the former part.

The non-chimeric end of INV4 local assembly 1:

AAGAGAAACTCAGAAATCTACGGAACAGGATAAAGGCAAGCGGGTAGCTGATGGCAGTC
CTGAACGAGAGGTACCCCCTGGACATAAGAAAGGCGTGTATGTCCGAATGATCATAAGC
GGACCGACTTTAGCGGGCCAATCTAGGAGGGCTATCAA

precisely matches the positive strand of scaffold_4 from 7515203rd bp to 7515358th bp.

INV4 local assembly 2 has one chimeric end:

CTTTGATTGGGTCGTCGGTTGGGGAGAGAGTGAAGGGGAGTGGGGTTAGGGTTTGGGGG
TCGAGTACTTGGTGGATGGTGAGGGAGCGGGTGTCCAGAGAGAGAGGGCCGGCGTTCAA
CCTTAATGAATAGGGTACACCGAAGG

in which the first 114 bases (black) match the negative strand of scaffold_4 from 3540530th bp to 3540417th bp, and the last 30 bases (red) match the positive strand of scaffold_4 from 7515406th bp to 7515435th bp.

The non-chimeric end of INV4 local assembly 2:

CAATAGTTGATCCAAGGCGCTTTTGAAAATGATGTCAACGGAGCTGCCGGTATCTACGA
GAGTTCTTGCTACCTTACTAGTGACAACCTTCAACGTAATTATGAAGGCATCGTCGTGG
GGGTACGAGATGCCCTCCCGAT

Includes 140 bases of which the first 136 match the negative strand of scaffold_4 from 7515614th bp to 7515479th bp, while the last 4 (CGAT) do not.

Real-Time PCR Validated Three Chromosome Rearrangements

For predicted large deletion DEL2 and large inversion INV4, the PCR results showed the expected novel products present in all five samples of Tango leaf DNA and none of the five samples of W. Murcott leaf DNA. For predicted translocation TRA1/3, Tango samples amplified the predicted product but a lesser amount of the same product was present in most W. Murcott leaf DNA samples.

DEL2:

Two sets of primers for validation of DEL2 local assembly 1 were designed according to the chimeric end, amplifying a segment spanning the apparent deletion breakpoints:

1. Forward primer NB81 and reverse primer NB82 (Figure 3.12). The melt peaks of the qPCR products showed that all five Tango DNAs and five W. Murcott DNAs samples amplified with MDH primers, and NTC did not amplify. With primers NB81 and NB82, Tango DNAs produced a single product while W. Murcott DNAs did not produce any product, the same as in NTC (Figure 3.5 upper panel).

2. Forward primer NB87 and reverse primer NB88 (Figures 3.12). The melt peaks of the qPCR products showed that all five Tango DNAs and five W. Murcott DNAs amplified with MDH primers, and NTC did not amplify. With primers NB87 and NB88, Tango DNAs produced a single product while W. Murcott DNAs did not produce any product, the same as in NTC. One W. Murcott DNA had a small peak at around 81.5 ℃, which might be due to contamination with MDH primers (Figure 3.5 middle panel).

Forward primer NB93 and reverse primer NB94 for validation of DEL2 local assembly 2 were designed according to the chimeric end, amplifying the join of the immediate upstream sequence of 12563421 bp and the immediate downstream sequence of 6572073 bp on scaffold_2 of the reference genome (Figure 3.12). The melt peaks of the qPCR products showed that all five Tango DNAs and five W. Murcott DNAs amplified with MDH primers, and NTC did not amplify. With primers NB93 and NB94, Tango DNAs

produced a single product while W. Murcott DNAs did not produce any product, the same as in NTC (Figure 3.5 lower panel).

TRA1/3:

Forward primer NB13 and reverse primer NB14 for validation of this TRA1/3 local assembly were designed according to the two non-chimeric ends of the local assembly respectively, amplifying a segment spanning the breakpoint. The melt peaks of qPCR showed that all five Tango DNAs and five W. Murcott DNAs amplified with MDH primers, and NTC did not amplify. With primers NB13 and NB14, Tango DNAs produced a single product and W. Murcott DNAs produced a product with the same melt temperature as Tango but with a weak signal, which indicates this may not be a simple translocation (Figure 3.6).

INV4:

Forward primer NB143 and reverse primer NB144 for validation of INV4 local assembly 1 were designed according to the chimeric end, amplifying a segment spanning the potential breakpoint (Figure 3.13). The melt peaks of qPCR showed that all five Tango DNAs and five W. Murcott DNAs amplified with MDH primers, and NTC did not amplify. With primers NB143 and NB144, Tango DNAs produced a single product while W. Murcott DNAs did not produce any product, the same as in NTC. One W. Murcott DNA had a small peak at around 83.5 ℃, which might be due to contamination (Figure 3.7 upper panel).

Two sets of primers were designed for validation of INV4 local assembly 2:

1. Forward primer NB115 and reverse primer NB116 were designed according to the chimeric end, amplifying a segment spanning the potential breakpoint (Figure 3.13). The qPCR melt peaks showed that all five Tango DNAs and five W. Murcott DNAs amplified with MDH primers, and NTC did not amplify. With primers NB115 and NB116, Tango DNAs produced a single product while W. Murcott DNAs did not produce any product, the same as in NTC. One W. Murcott DNA had a small peak at around 79 ℃, which might be due to contamination (Figure 3.7 middle panel).

2. Forward primer NB153 and reverse primer NB154 designed according to both chimeric and non-chimeric ends, amplifying a segment spanning the potential breakpoint (Figure 3.13). The qPCR melt peak showed that all five Tango DNAs and five W. Murcott DNAs amplified with MDH primers, and NTC did not amplify. With primers NB153 and NB154, Tango DNAs produced a single product while W. Murcott DNAs did not produce any product, the same as in NTC (Figure 3.7 lower panel).

The results above showed that there was a clear difference between W. Murcott and Tango leaf DNAs with the three sets of primers for both the large deletion and the large inversion, in which all five Tango DNA samples amplified while none of the five W. Murcott DNA samples did, indicating the sequences of these amplicons only exist in Tango. In contrast, the translocation did not show clear and easily interpreted differences between W. Murcott and Tango although the signal strength of W. Murcott was much weaker than that of Tango.

Sanger Sequencing of Regions Flanking Breakpoint

Deletion 2 – Breakpoint 1 Sanger Sequencing Result

TTCTTCATTTTTGAATTTGAGTGGCCAACAAATAAAATTATTTGAAAACTACTCTATCC
TGAGAGTTATTGGTAGATGTGTGATTGCTAACCACCAGAAAAAATAACAAAAAAGAAAT
TTAATTGAGGAAAATCAAATATCTATAAAGATTTGTACCAATCATGTACACAAATCAAA
CTCCCTCAGTGTTCTAAAGTGTAAAGGCCAGAGAAGACCTTGACTCACTATTGTGTTTG
AAAACGAGTGTTGCTATTCAACGCGGGAAGAAGATGATTATGGGCCTGCTCATGGTGTC
CTTTTCCTGGGTTAAAGTGTCTA**TAT**TTTAAATTTTGAAGGGATTGATGGAGTAGGACT
TGGACTGGTTAGGCACCATTACTCGCTTGGAAAGTAAGTAATGTCCAAGATTTGAGTGC
TTTATACCATCAATCATTGATAATCAGAACTGTTGAGAAGTTTTCGAAATTGAAGCCAT
CTTGATGTGGACATCCAATGATTCACCATGTATGTCTGAGTAATTTATCGATTTTTTTA
TTAATGAAGGATTTTTGATATATGGTGCTTGCATATTCCTGTCAAGGATCGAACGTCGT
TTGCAGGTATCGTTATTCTTTGTGGTGG

BLAST of the above sequence against the *Citrus clementina* reference genome showed that from 1st bp to 321st bp can be aligned to scaffold_2 from 6571746th bp to 6572075th bp, and from 319th bp to 618th bp can be aligned to scaffold_2 from 12563421st bp to 12563720th bp. There is a TAT trinucleotide at the breakpoint which can be aligned to either end of the sequences flanking the deletion (Figure 3.8).

Deletion 2 – Breakpoint 2 Sanger Sequencing Result

CACCCTATCAAAGTGAAGGCATTTGTATTATGTATGTTTGATTCATAAAGGCATTGCAC
ATATGAATGGTAATGGACTCCTTGGTTTATAATGCTGTAGTATGTTATATATGTTAAGT

CATTATAGCTCAATTGTTGATTTGCCAATTTAGAATTTATTTGTTGCTGGAGTATTTTC

TTTTTTCAAATGGTTTACTTATAGCTTTGAAGCTCTGAGTGGAAATAGCATGAATTATT

CATTTTTTTT**T**ATGAACTTTATAAGAACTTGCAGAATCCAATCAACCGAAAATTTACAAT

CATGGATCCGACTTCCTTCATTTTCAGATTGGCGGCCCTTGTATGGGCTGTAATATTTA

CATGCGATAACGTTGCTG

BLAST of the above sequence against the *Citrus clementina* reference genome showed that from 16th bp to 246th bp can be aligned to scaffold_2 from 12563183th bp to 12563420th bp, and from 246th bp to 372nd bp can be aligned to scaffold_2 from 6572073rd bp to 6572199th bp, where the T at 246th bp can be aligned to either end of the sequences within the deletion (Figure 3.9). The nucleotides that were not aligned had low quality Sanger sequences.

Inversion 4 – Breakpoint 1 Sanger Sequencing Result

CCCACCAAGTTCATGATTCCTTAAATTCCAATCTCCTCTCTCAAACAATCAACCCCCGA

GAAAACACTCACTCAGTCCTCCCTCACTGACCGGGTTAACTCAATTTCCACCATAACAA

CAATGGCACCGATCGACCCGCACTCGTTCACCGAGTCAACTCACCCGCTGACCACACAC

ATCTCCCTCTCTCTCTACTTCGACTTCTCTTCATCCACCATCCACGCCGCCGCCATCCT

TACCCTCGCCTCC**CCA**ATATTGGTGGTTATCGACAACGATCTACCATAGCCTTTGATAG

CCCTCCTAGATTGGCCCGCTAAAGTCGGTCCGCTTATGATCATTCGGACATACACGCCT

TTCTTATGTCCAGGGGGTACCTCTCGTTCAGGACTGCAATCAGCTACCCGCTTGCCATA

ATCCTGTTCCGTAAATTTCGGAGTTTCTCTGATGCCAGCAACAAATTCTTTTAGGTACC

CGCTGAGCATAAAAGATTCAATTTCCTCTCGCAAAGTGAAGCACA

BLAST of the above sequence against *Citrus clementina* reference genome showed that from 1st bp to 252nd bp can be aligned to scaffold_4 from 3540164th bp to 3540415 th bp, and from 250th bp to 516th bp can be aligned to the reverse complement of scaffold_4 from 7515131st bp to 7515397th bp. There is a CCA trinucleotide at the breakpoint which can be aligned to either end of the inversion (Figure 3.10).

Inversion 4 – Breakpoint 2 Sanger Sequencing Result

AGGTGGAGAAGACGATGAGTACAGAGGAGTGATCGGAGAGCGTGACGATGAGGTGGCGG
CCTTTGATTGGGTCGTCGGTTGGGGAGAGAGTGAAGGGGAGTGGGGTTAGGGTTTGGGG
GTCGAGTACTTGGTGGATGGTGAGGGAGCGGGTGTCCAGAGAGAGAGGGCCGGCGTTCA
ACCTTAATGAATAGGGTACACCGAAGGTACCCCACCATCCCCCGCCTACCCTTTTTACA
GAAAAAGATGCGGAGGGCATCTCGTACCCCCACGACGATGCCTTCATAATTACGTTGAA
GGTTGTCACTAGTAAGGTAGCAAAGCAACCATCTACCGGCAGCTCCGTCGACATCATTT
ACAAAAGCGCCTGGATCAACTATTGATGACTCCCCAAGGATCCCACTG

BLAST of the above sequence against the *Citrus clementina* reference genome showed that from 1st bp to 174th bp can be aligned to the reverse complement of scaffold_4 from 3540417th bp to 3540590th bp, and from 175th bp to 318th bp can be aligned to scaffold_4 from 7515406th bp to 7515549th bp (Figure 3.11).

Discussion

Chromosome rearrangements have been found in other citrus varieties. Both heterozygous translocation and inversion in *Citrus assamensis* could be the first record of chromosome rearrangement heterozygosity in citrus (NAITHANI AND RAGHUVANSHI, 1958; RAGHUVANSHI, 1962). A quadrivalent was observed in Lue Gim Gong orange, which resulted from a reciprocal translocation and caused gametic sterility (IWAMASA, 1970). In 'Valencia' sweet orange, a quadrivalent was observed in meiosis of pollen mother cells and confirmed to be a heterozygous chromosomal reciprocal translocation (LAN et al., 2016). The partial pollen sterility of 'Mexican' lime is caused by a heterozygous inversion and low temperatures, which reduces chromosome pairing (IWAMASA, 1970; OLLITRAULT et al., 2007). Induced chromosome rearrangement has also been widely used in other species. The chromosomal translocation lines of watermelon induced by ^{60}Co γ radiation not only produce few seeds but also overcome the low germination and poor seedling establishment of triploid seedless watermelon (WU et al., 2013). Translocations were also confirmed after gamma irradiation of *Nigella damascena* L., where a quadrivalent was present in heterozygotes and lead to defective female gametogenesis and seed setting failure (ANINDITA AND DATTA, 2006).

The chromosome rearrangements can be predicted with NGS reads. However there are some types of false positives in NGS and analysis: machine-sequencing errors, incorrect local alignment of individual reads and discordant alignment of pairs, and failure to detect

alleles that are different from the reference sequence (MEYERSON et al., 2010). Machine-sequencing errors are stochastic errors, can be overcome by high-level over-sampling of DNA with more stringent statistical thresholds, and the failure of detection of the alleles different from the reference sequence can be overcome by sufficient coverage (MEYERSON et al., 2010). The Illumina data of W. Murcott and Tango used here are of about 25X coverage, which is a fair coverage. However due to the incompleteness of the *Citrus clementina* assembly genome (approximately 294 Mb for the nine chromosome pseudomolecules), reads from outside of the sequence of the pseudochromosomes could not be mapped or mapped to incorrect positions, which could be a major reason for inaccurate structural variation analysis (WU et al., 2013). Also, the existence of duplication in the reference genome could be another reason for the difficulty of structural variation analysis, in which reads might not be mapped to their correct positions.

The potential for false discovery of structural variations that are based on discordant paired-end read mapping can be overcome by recently developed methods for direct mapping on single molecules such as Bionano optical genome mapping, in which long DNA molecules can be directly imaged in its native state (STAŇKOVÁ et al., 2016). Chromosome rearrangements can be verified if mapping positions are different in mutant from the parent.

Copy Number Variant Detection of Large Segments

If a chromosomal segment that is sufficiently large is deleted or duplicated compared to the reference genome coverage of that segment will be decreased or increased, given the

condition that the sampling of all the reads was evenly distributed. In our study, the *Citrus clementina* reference genome is incomplete and the Tango genome is expected to be more similar to that of W. Murcott than to the reference. Therefore, it is best to compare the coverage between Tango and W. Murcott along the reference genome to reduce the influence of the differences between the NGS data and the reference genome.

If any large chromosomal segment which has a single copy in the haploid reference genome is deleted in Tango compared to W. Murcott, the coverage of that region should reduce to half for a heterozygous deletion or zero for a homozygous deletion. Duplications relative to the reference should increase fold coverage in proportion to copy number. The result showed the log2 ratio of coverage of Tango versus W. Murcott of all nine chromosomes varied between -0.3 and 0.3, without a large region whose value was close to -1 or 1. The largest difference was around $2^{0.3} = 1.23$ fold and regions with the largest differences were not contiguous. This may indicate there was no clear copy number variation larger than 50 kb between Tango and W. Murcott along all nine chromosomes under the condition that the read sampling is uniform, or Tango had one or a few copies deleted or duplicated when W. Murcott already had many copies of that region, which makes the difference not significant. The other possibility is chimerism, where either W. Murcott or Tango (the more likely) is chimeric for a deleted or duplicated region while the other is not, but only a small portion of the sampled cells of the mutant carry the copy number variant.

The LOH analysis of the SNP array also provided evidence that Tango does not carry any large heterozygous deletions. A heterozygous deletion that is large (>1 Mb) would present as a contiguous region with 100% or close to 100% LOH, and without heterozygous SNPs

detected in Tango. However, the percentage of LOH through all chromosomes was relatively low, except a few single non-consecutive windows, where the maximum LOH percentage is no more than 50%. The percentage of LOH at 19 Mb of chromosome 2 could not be estimated because the array contained no SNPs heterozygous for W. Murcott in this region. The density plot of all the SNP markers also showed no dense LOH data points in regions where heterozygous markers were at low density. The above result indicates that there were only SNPs instead of large heterozygous deletions present in Tango compared to W. Murcott or the deletions were likely too small to cause asynapsis during meiosis of Tango or many aneuploid gametes. Over the entire genome, 96.7% of 1 Mb intervals had at least 10 heterozygous SNPs in W. Murcott, indicating the array is a useful platform to identify most large LOH regions.

In addition, based on the SNP array data, we observed some adjacent LOH position clusters, which may indicate small deletions. On chromosome 2, there is a cluster of three adjacent LOH positions at 22717498th bp, 22913950th bp and 22955060th bp, flanked by two closest non-LOH positions at 22690908th bp and 22991472nd bp. And on chromosome 5, there is another cluster of three adjacent LOH positions at 34809444th bp, 34946092nd bp and 35024615th bp, flanked by two closest non-LOH positions at 34710901st bp and 35044945th bp. No other cluster with equal or more than three adjacent LOH positions was observed. Based on the positions of closest flanking non-LOH of the clusters, the size of the possible deletion on chromosome 2 could be between 238 kb to 301 kb, and the possible deletion on chromosome 5 could be between 215 kb to 334 kb, neither of which are larger than 1 Mb.

Combining the results above, it is possible that no segment of DNA that is sufficiently large to cause abnormal chromosomal pairing during meiosis is duplicated (detected as coverage increase) or deleted (detected as coverage decrease and LOH) in Tango compared to W. Murcott.

Chromosome Rearrangement Identification and Breakpoint Determination

Both a single balanced translocation and a single balanced inversion with a crossover event between the inverted and normal homologous chromosome could cause half of the derived gametes to be aneuploid resulting in potentially non-functional pollen or inviable zygotes. Based on the viability and germination rate of Tango pollen of previous studies (ROOSE AND WILLIAMS, 2007), and observations of meiosis in Tango there could only be a few structural variations that affect meiosis (CROWLEY, 2011).

The output of novoBreak program identified 29 potential structural variations, which is more than expected. After removal of small structural variations which are unlikely to cause abnormal chromosomal synapsis and combining predicted rearrangements with similar breakpoints, only nine structural variations remained for verification. The following PCR gave positive results for three out of the nine structural variations, each of which had at least two supporting Illumina paired-end reads for the corresponding structural variation. The others were not validated, either because no reads that span two chromosomes or two positions on a chromosome could be found in the NGS data, or the read sequences were not suitable for primer design and were not tested further.

The small proportion of structural variations that could be validated may be due to the following reasons: 1. The reference *Citrus clementina* genome is not complete or contains assembly errors. Those reads that came from chromosomal regions that are not in the reference genome were mistakenly aligned to similar regions or abandoned due to the low similarity, which could result in both false positive and false negative structural variation predictions. 2. Although the varieties are closely related, there are still some slight differences between the *Citrus clementina* mandarin genome and the W. Murcott/Tango mandarin genome. Differences could result in the misalignment of reads from the sequences that are present in W. Murcott or/and Tango but not in Clementine genome, which also cause false positive and false negative prediction. 3. Due to the potential chimerism of Tango, cells from some tissue or cell layers may carry a structural variation while cells from others do not. The read alignment program cannot distinguish reads with structural variation from sequencing errors and reads spanning chromosome rearrangement breakpoints may have been classified as reads of low mapping quality and eventually abandoned, which could result in the underestimation or omission of some structural variations that objectively exist.

For large deletion DEL2, two local read assemblies were verified. The first local read assembly is mapped as flanking the "deleted" segment, that is, spanning the two breakpoints. The sequencing result supported the prediction that chromosome segments upstream at between 6572073rd bp – 6572075th bp and downstream at between 12563421st bp – 12563423rd bp of chromosome 2 joined together. The second local read assembly showed the two ends of the "deleted" segment between the two breakpoints, which are

downstream at between 6572073rd bp – 6572075th bp and upstream at between 12563421st bp – 12563423rd bp, were joined together, which was also supported by sequencing. The BLAST result of Sanger sequencing of the PCR product Deletion 2 - breakpoint 1 at the junction point showed the sequence has a trinucleotide TAT at the junction point which can be either aligned to 6572073rd bp – 6572075th bp or 12563421st bp – 12563423rd bp (red boxes in Figure 3.8). Similarly, there is a T at the junction point which can be aligned to either 12563420th bp or 6572073rd bp of the PCR product Deletion 2 - breakpoint 2 (red boxes in Figure 3.9). Since the T at 6572073rd bp may not be shared simultaneously by two segments of DNA on either side of breakpoint, it can be concluded that the sequence upstream of 6572073rd bp (included) is joined to that downstream of 12563421st bp (included), and the sequence downstream of 6572074th bp (included) is joined to that upstream of 12563420th bp (included), with another potential DSB between two breakpoints. Due to the presence of sequence reads and PCR products of the sequence around the junction point, as well as no decreased read coverage observed in CNV analysis or uniformly changed heterozygosity of SNP observed in LOH analysis within this 6 Mb region, this intervening chromosomal segment was in fact not deleted but perhaps translocated to somewhere else in the genome. In addition, based on merely the mapping positions and orientation of reads that are involved between the two breakpoints, it appears to be a tandem duplication of a 6 Mb segment which starts from sequence immediately downstream of 6572074th bp and ends with sequence immediately upstream of 12563420th bp. However, since no copy number increase was observed through all chromosomes, the 6 Mb segment as a whole is likely not a tandem duplication. In such cases, there should be

another breakpoint within this 6 Mb segment which is yet discovered. To investigate whether it is a tandem repeat, real-time PCR can be employed and copy number can be compared with an internal control. The other possibility is the occurrence of an intrachromosomal translocation, where a segment of DNA upstream of 6572073rd bp (included) is switched with a segment of DNA upstream of 12563420th bp (included), or a segment of DNA downstream of 6572074rd bp (included) is switched with a segment of DNA downstream of 12563421st bp (included), either case of which includes another two undiscovered breakpoints (Figure 3.12). Therefore, as mentioned above, the location and sequence of another breakpoint or breakpoints must be discovered to determine the location and rearrangement of the 6 Mb segment. In the Illumina sequence data, two SNPs were found near 12563420th bp: a G at 12563219th bp is linked to a T at 12563281st bp, and a C at 12563219th bp is linked to a G at 12563281st bp, representing a GT haplotype and a CG haplotype in both W. Murcott and Tango. In the Tango read alignment, there are 12 reads with G and 22 reads with C at the 12563219th bp, and 18 reads with T and 15 reads with G at the 12563281st bp. All five chimeric reads involved in DEL2 carry G and T alleles at these positions, which indicates the GT haplotype was affected by DEL2. Near the other breakpoint at 6572073rd bp, many reads presenting various haplotypes were aligned, resulting in the local depth of coverage around this breakpoint of over 100X. The excess depth of coverage may be because of reads carrying different SNPs were misaligned due to sequence similarity, an indication that additional copies of a similar region likely exist in W. Murcott and Tango.

For translocation TRA1/3, no strong evidence supported a translocation between chromosome 1 and chromosome 3. Firstly, the reads which mapped to the two chromosomes involved in the predicted translocation were all from one strand of DNA, and no reads from the other strand were discovered from NGS data. This result may due to the sampling bias if a translocation was present but reads from the other strand were missed, or due to the incorrect mapping of reads that do not belong to the position. Secondly, from the investigation of the breakpoints on chromosome 1 and chromosome 3, only reads with one end completely mapped to chromosome 1 and the other end partially mapped to chromosome 3 (breakpoint at one end of a read-pair) were found, while no reads with one end completely mapped to chromosome 3 and partially mapped to chromosome 1 (breakpoint at the other end of a pair of read), or both ends were completely mapped to each corresponding chromosome (breakpoint between the two ends of a pair of read) were found. Thirdly, in the case of a reciprocal translocation, we expect reads that map to left side of the breakpoint on Chromosome A and to right side of the breakpoint on Chromosome B, and also reads that map to the right side of the breakpoint on Chromosome A and to the left side of the breakpoint on Chromosome B, respectively. In our study, only reads and read pairs spanning the breakpoints that have one end mapped to lower coordinate of chromosome 1 (upstream of 27525574th bp) and the other end mapped to higher coordinate of chromosome 3 (downstream of 31452544th bp) were found, while the reciprocal breakpoint-spanning reads or read pairs, which were supposed to have one end mapped to higher coordinate of chromosome 1 (downstream of 27525574th bp) and the other end mapped to lower coordinate of chromosome 3 (upstream of 31452544th bp), were

not found. The case that only unbalanced reads were found for this predicted translocation indicating an unbalanced translocation which should also have large deletions observed. However, the LOH analysis of SNP array data also showed that a large segment with LOH is unlikely. In addition, BLAST results for the sequence assembly of the two ends of the breakpoint-spanning reads show that one end can be uniquely aligned to chromosome 1 at the corresponding position (lower coordinate of breakpoint), while the other end consists of two parts: one part is 44 bp of low complexity sequence with stretch of A's and T's (AAATA**R**AAATTAAAAAAAAAAATTTTTTTGTTTTTTTTGTTTTT) that could not be aligned specifically and the other part can be aligned to multiple positions in the genome. The inability to uniquely align this sequence assembly could result from either the insertion of a replicate as a result of transposons in Tango genome, or the Illumina reads being incorrectly aligned. Fourthly, the real-time PCR amplification with primers flanking the breakpoint did not show a significant difference between Tango and W. Murcott, since most W. Murcott samples had a product with the same melt temperature as Tango but with a weaker signal, which means the sequence of the amplicon may present in W. Murcott genome with fewer copies than that in Tango genome, indicating this is less likely a simple translocation, although it seems to reflect a difference between W. Murcott and Tango.

For inversion INV4, two assemblies were verified, each included a pair of assembled read ends: one assembly had one end inside the inverted 4 Mb segment without including the breakpoint while the other end spanned the breakpoint at around 3.5 Mb chromosomal position. The other assembly had the one end outside the 4 Mb inverted segment without including the breakpoint while the other end spanned the breakpoint at around 7.5 Mb

chromosomal position. BLAST of the Illumina sequence and Sanger sequences at the two junction points of the predicted inverted segment confirmed this 4 Mb segment is inverted in respect to the reference genome. The BLAST result of Sanger sequences of the PCR product from Inversion 4-1 (red boxes Figure 3.10) at the junction point showed the sequence has a trinucleotide CCA at the junction point that can be either aligned to 3540413^{th} bp – 3540415^{th} bp or 7515395^{th} bp – 7515397^{th} bp on reference genome coordinate. It can be concluded that the breakpoints are between 3540412^{th} bp – 3540416^{th} bp and 7515394^{th} bp – 7515405^{th} bp, where the CCA trinucleotide can be assigned to either end. A few nucleotides connecting 3540413^{th} bp – 3540415^{th} bp to 3540417^{th} bp and connecting 7515395^{th} bp – 7515397^{th} bp to 7515406^{th} bp at the two breakpoints were lost compared to the reference genome, possibly due to the DSB repair process (Figure 3.13). According to the previous study on centromere mapping, INV4 is likely to be a paracentric inversion (ALEZA et al., 2015). By analyzing the read alignment near the breakpoints, two SNPs were observed. There is an A at 7515335^{th} bp linked with a T at 7515520^{th} bp and a G at 7515335^{th} bp linked with a C at 7515520^{th} bp in non-chimeric reads in both W. Murcott and Tango, representing an AT haplotype and a GC haplotype in both W. Murcott and Tango. The SNP frequency in Tango showed that there are 13 reads with A and 16 reads with G at 7515335^{th} bp, and 10 reads with T and 11 reads with C at 7515520^{th} bp. Three of all the reads involved in INV4 inversion cover both SNPs, where there is a G at 7515335^{th} bp and a C at 7515520^{th} bp, indicating the GC allele of Tango was inverted. No SNPs were observed near the other breakpoint of INV4.

According to the genome browser on Citrus Genome Database, on the Clementine genome, one of the two breakpoints of DEL2 at 12563420 on chromosome 2 is within the gene Ciclev10014550m.g, which is predicted to encode an acyltransferase-like protein At3g26840. One of the two breakpoints of INV4 at 3540415 on chromosome 4 is within the gene Ciclev10030974m.g, which is predicted to be a leukotriene A-4 hydrolase homolog. No clear evidence indicates that these genes are related to meiosis or gamete development, while the possibility cannot be excluded. The other breakpoints of three chromosome rearrangements were not predicted to be within genes (Table 3.5).

In summary, the predicted translocation TRA1/3 could be an insertion of a small segment caused by transposons, the predicted large deletion DEL2 had the "deleted" region relocated in somewhere else in the genome, and the predicted large inversion INV4 matched the experimental result. None of the above would cause copy number variation or LOH, which matches the analysis of coverage and LOH from the SNP array, and the inversion was sufficiently large to cause abnormal synapsis and disjunction during metaphase and thus result in infertile pollen. The breakpoint positions were localized within a few nucleotides. Greater precision was not possible perhaps because exonuclease activity removes some nucleotides during the DSB repair, or because there are differences between the W. Murcott genome and *Citrus clementina* genome.

References

Aleza, P., J. Cuenca, M. Hernandez, J. Juarez, L. Navarro, and P. Ollitrault, 2015. Genetic Mapping of Centromeres in the Nine Citrus Clementina Chromosomes Using Half-Tetrad Analysis and Recombination Patterns in Unreduced and Haploid Gametes. *BMC Plant Biol* 15:80.

Anindita, G. and A.K. Datta, 2006. Gamma-Rays Induced Reciprocal Translocation in Nigella Damascena L. (Love-in-a-Mist). *Caryologia* 59:31-36.

Barkley, N.A., K.D.C. Chamberlin, M.L. Wang, and R.N. Pittman, 2010. Development of a Real-Time PCR Genotyping Assay to Identify High Oleic Acid Peanuts (Arachis Hypogaea L.). *Molecular Breeding* 25:541-548.

Chong, Z., J. Ruan, M. Gao, W. Zhou, T. Chen, X. Fan, L. Ding, A.Y. Lee, P. Boutros, J. Chen, and K. Chen, 2017. Novobreak: Local Assembly for Breakpoint Detection in Cancer Genomes. *Nat Methods* 14:65-67.

Crowley, J.R., 2011. A Molecular Genetic Approach to Evaluate a Novel Seedless Phenotype Found in Tango, a New Variety of Mandarin Developed from Gamma-Irradiated W. Murcott. Ph.D. Dissertation, University of California, Riverside.

Gaeta, R.T. and J.C. Pires, 2010. Homoeologous Recombination in Allopolyploids: The Polyploid Ratchet. *New Phytologist* 186:18-28.

Gambetta, G., A. Gravina, C. Fasiolo, C. Fornero, S. Galiger, C. Inzaurralde, and F. Rey, 2013. Self-Incompatibility, Parthenocarpy and Reduction of Seed Presence in 'Afourer' Mandarin. *Scientia Horticulturae* 164:183-188.

Gill, B.S., C.R. Burnham, G.R. Stringam, J.T. Stout, and W.H. Weinheimer, 1980. Cytogenetic Analysis of Chromosomal Translocations in the Tomato - Preferential Breakage in Heterochromatin. *Canadian Journal of Genetics and Cytology* 22:333-341.

Goodwin, S., J.D. McPherson, and W.R. McCombie, 2016. Coming of Age: Ten Years of Next-Generation Sequencing Technologies. *Nat Rev Genet* 17:333-351.

Gorbunova, V. and A.A. Levy, 1999. How Plants Make Ends Meet: DNA Double-Strand Break Repair. *Trends in Plant Science* 4:263-269.

Gunderson, K.L., F.J. Steemers, G. Lee, L.G. Mendoza, and M.S. Chee, 2005. A Genome-Wide Scalable SNP Genotyping Assay Using Microarray Technology. *Nature Genetics* 37:549-554.

Hui, L., T. DelMonte, and K. Ranade, 2008. Genotyping Using the Taqman Assay. *Current Protocols in Human Genetics*:2.10. 11-12.10. 18.

Iwamasa, M., 1970. Chromosome aberrations in citrus in relation to sterility and seedlessness, 1st International Citrus Symposium, Riverside 1968.

Jancey, R. and D. Walden, 1972. Analysis of Pattern in Distribution of Breakage Points in the Chromosomes of Zea Mays L. And D. Melanogaster Meigen. *Canadian Journal of Genetics and Cytology* 14:429-442.

Khush, G.S. and C.M. Rick, 1968. Cytogenetic Analysis of the Tomato Genome by Means of Induced Deficiencies. *Chromosoma* 23:452-484.

Kidd, J.M., G.M. Cooper, W.F. Donahue, H.S. Hayden, N. Sampas, T. Graves, N. Hansen, B. Teague, C. Alkan, F. Antonacci, E. Haugen, T. Zerr, N.A. Yamada, P. Tsang, T.L. Newman, E. Tuzun, Z. Cheng, H.M. Ebling, N. Tusneem, R. David, W. Gillett, K.A. Phelps, M. Weaver, D. Saranga, A. Brand, W. Tao, E. Gustafson, K. McKernan, L. Chen, M. Malig, J.D. Smith, J.M. Korn, S.A. McCarroll, D.A. Altshuler, D.A. Peiffer, M. Dorschner, J. Stamatoyannopoulos, D. Schwartz, D.A. Nickerson, J.C. Mullikin, R.K. Wilson, L. Bruhn, M.V. Olson, R. Kaul, D.R. Smith, and E.E. Eichler, 2008. Mapping and Sequencing of Structural Variation from Eight Human Genomes. *Nature* 453:56-64.

Kronenberg, A. and J.B. Little, 1989. Molecular Characterization of Thymidine Kinase Mutants of Human Cells Induced by Densely Ionizing Radiation. *Mutat Res* 211:215-224.

Lan, H., C.L. Chen, Y. Miao, C.X. Yu, W.W. Guo, Q. Xu, and X.X. Deng, 2016. Fragile Sites of 'Valencia' Sweet Orange (Citrus sinensis) Chromosomes Are Related with Active 45s rDNA. *PLoS One* 11:e0151512.

Lee, M. and R.L. Phillips, 1988. The Chromosomal Basis of Somaclonal Variation. *Annual Review of Plant Physiology and Plant Molecular Biology* 39:413-437.

Longley, A.E., 1961. Breakage Points for Four Corn Translocation Series and Other Corn Chromosome Aberrations Maintained at the California Institute of Technology. *ARS 34; 16*.

Mei, M., H. Deng, Y. Lu, C. Zhuang, Z. Liu, Q. Qiu, Y. Qiu, and T.C. Yang, 1994. Mutagenic Effects of Heavy Ion Radiation in Plants. *Adv Space Res* 14:363-372.

Meyerson, M., S. Gabriel, and G. Getz, 2010. Advances in Understanding Cancer Genomes through Second-Generation Sequencing. *Nature Reviews Genetics* 11:685-696.

Naithani, S. and S. Raghuvanshi, 1958. A Preliminary Meiotic Study in Citrus Assamensis—a Structural Hybrid. *Naturwissenschaften* 45:95-95.

Nordfors, L., M. Jansson, G. Sandberg, C. Lavebratt, S. Sengul, M. Schalling, and P. Arner, 2002. Large- Scale Genotyping of Single Nucleotide Polymorphisms by Pyrosequencing™ and Validation against the 5′ Nuclease (Taqman®) Assay. *Human Mutation* 19:395-401.

Ollitrault, P., Y. Froelicher, D. Dambier, F. Luro, and M. Yamamoto, 2007. Seedlessness and Ploidy Manipulation. *Citrus Genetics, Breeding and Biotechnology*:197-218.

Parker, J.S., 1987. Increased Chiasma Frequency as a Result of Chromosome Rearrangement. *Heredity* 58:87-94.

Pipiras, E., A. Coquelle, A. Bieth, and M. Debatisse, 1998. Interstitial Deletions and Intrachromosomal Amplification Initiated from a Double-Strand Break Targeted to a Mammalian Chromosome. *EMBO Journal* 17:325-333.

Puchta, H., 2005. The Repair of Double-Strand Breaks in Plants: Mechanisms and Consequences for Genome Evolution. *J Exp Bot* 56:1-14.

Raghuvanshi, S., 1962. Cytogenetical Studies in Genus Citrus Citrus Assamensis. *Caryologia* 15:143-149.

Rieger, R., A. Michaelis, I. Schubert, P. Döbel, and H.-W. Jank, 1975. Non-Random Intrachromosomal Distribution of Chromatid Aberrations Induced by X-Rays, Alkylating Agents and Ethanol in Vicia Faba. *Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis* 27:69-79.

Rieseberg, L.H., 2001. Chromosome rearrangements and Speciation. *Trends Ecol Evol* 16:351-358.

Roose, M.L. and T.E. Williams, 2007. Mandarin Tree Named 'Tango'. U.S. Patent No. PP17,863.

Rowley, J.D., 1982. Identification of the Constant Chromosome Regions Involved in Human Hematologic Malignant Disease. *Science* 216:749-751.

Roy, S., 2014. Maintenance of Genome Stability in Plants: Repairing DNA Double Strand Breaks and Chromatin Structure Stability. *Frontiers in Plant Science* 5.

Sankaranarayanan, K., R. Taleei, S. Rahmanian, and H. Nikjoo, 2013. Ionizing Radiation and Genetic Risks. Xvii. Formation Mechanisms Underlying Naturally Occurring DNA Deletions in the Human Genome and Their Potential Relevance for Bridging the Gap between Induced DNA Double-Strand Breaks and Deletions in Irradiated Germ Cells. *Mutation Research-Reviews in Mutation Research* 753:114-130.

Saxena, R.K., D. Edwards, and R.K. Varshney, 2014. Structural Variations in Plant Genomes. *Brief Funct Genomics* 13:296-307.

Staňková, H., A.R. Hastie, S. Chan, J. Vrána, Z. Tulpová, M. Kubaláková, P. Visendi, S. Hayashi, M. Luo, and J. Batley, 2016. Bionano Genome Mapping of Individual Chromosomes Supports Physical Mapping and Sequence Assembly in Complex Plant Genomes. *Plant Biotechnology Journal* 14:1523-1531.

Sung, P. and H. Klein, 2006. Mechanism of Homologous Recombination: Mediators and Helicases Take on Regulatory Functions. *Nat Rev Mol Cell Biol* 7:739-750.

Szostak, J.W., T.L. Orr-Weaver, R.J. Rothstein, and F.W. Stahl, 1983. The Double-Strand-Break Repair Model for Recombination. *Cell* 33:25-35.

Thacker, J., 1986. The Nature of Mutants Induced by Ionizing-Radiation in Cultured Hamster-Cells .3. Molecular Characterization of Hprt-Deficient Mutants Induced by Gamma-Rays or Alpha-Particles Showing That the Majority Have Deletions of All or Part of the Hprt Gene. *Mutation Research* 160:267-275.

Weckselblatt, B. and M.K. Rudd, 2015. Human Structural Variation: Mechanisms of Chromosome Rearrangements. *Trends Genet* 31:587-599.

Wu, G.A., S. Prochnik, J. Jenkins, J. Salse, U. Hellsten, F. Murat, X. Perrier, M. Ruiz, S. Scalabrin, J. Terol, M.A. Takita, K. Labadie, J. Poulain, A. Couloux, K. Jabbari, F. Cattonaro, C. Del Fabbro, S. Pinosio, A. Zuccolo, J. Chapman, J. Grimwood, F.R. Tadeo, L.H. Estornell, J.V. Munoz-Sanz, V. Ibanez, A. Herrero-Ortega, P. Aleza, J. Perez-Perez, D. Ramon, D. Brunel, F. Luro, C. Chen, W.G. Farmerie, B. Desany, C. Kodira, M. Mohiuddin, T. Harkins, K. Fredrikson, P. Burns, A. Lomsadze, M. Borodovsky, G. Reforgiato, J. Freitas-Astua, F. Quetier, L. Navarro, M. Roose, P. Wincker, J. Schmutz, M. Morgante, M.A. Machado, M. Talon, O. Jaillon, P. Ollitrault, F. Gmitter, and D. Rokhsar, 2014. Sequencing of Diverse Mandarin, Pummelo and Orange Genomes Reveals Complex History of Admixture During Citrus Domestication. *Nature Biotechnology* 32:656-662.

WU, J.-y., Y.-h. WU, and X.-l. XIE, 2013. Breeding of Chromosome Translocation Line and F_1 Hybrid with Reduced Seed in the Fruit of Watermelon. *China Cucurbits and Vegetables* 1:003.

Zhu, X., Y. Xu, S. Yu, L. Lu, M. Ding, J. Cheng, G. Song, X. Gao, L. Yao, and D. Fan, 2014. An Efficient Genotyping Method for Genome-Modified Animals and Human Cells Generated with CRISPR/Cas9 System. *Scientific Reports* 4.

Table 3.1 Summary of Real-Time PCR Primers Amplifying Chimeric Sequences Formed by Putative Structural Variants in Tango

Structural variation type	Assembly	Primer name	Sequence	Reference genome hits
Large deletion DEL2	Large deletion local assembly 1	NB81	CCAAGTCCTACTCCATCAATC	scaffold_2: 12563459-12563439
		NB82	TTTCCTGGGTTAAAGTGTCTAT	scaffold_2: 6572052-6572073
	Large deletion local assembly 2	NB87	CCTGCTCATGGTGTCCTTT	scaffold_2: 6572035-6572053
		NB88	GTCCTACTCCATCAATCCCTTC	scaffold_2: 12563455-12563434
	Large deletion local assembly 3	NB93	GGTTGATTGGATTCTGCAAGT	scaffold_2: 6572109-6572089
		NB94	CTCTGAGTGGAAATAGCATGAATTA	scaffold_2: 12563384-12563408
Translocation TRA1/3	Translocation local assembly	NB13	GAGACAAGCACCCCCAATG	scaffold_1: 27525506-27525524
		NB14	TGAAAAGCCAATGATAAGAATTTG	scaffold_3: 31452462-31452485 [z]
Large inversion INV4	Large inversion local assembly 1	NB143	ACACATCTCCCTCTCTCTCTAC	scaffold_4: 3540337-3540358
		NB144	GCCAATCTAGGAGGGCTATC	scaffold_4: 7515337-7515356 [z]
	Large inversion local assembly 2 set 1	NB115	CGGGTGTCCAGAGAGAGA	scaffold_4: 3540444-3540427
		NB116	GGTGTACCCTATTCATTAAGGTTG	scaffold_4: 7515430-7515407
	Large inversion local assembly 2 set 2	NB153	CGGGTGTCCAGAGAGAGA	scaffold_4: 3540444-3540427
		NB154	ACTAGTGACAACCTTCAACGTAA	scaffold_4: 7515539-7515517

[z] Primer had multiple hits in BLAST against the reference genome.

Table 3.2 Summary of Conventional PCR Primers Amplifying Chimeric Sequences for Sanger Sequencing

Structural variation type	Primer name	Sequence	Reference genome hits
Deletion 2-1	nBDEL_LG_F	TTCTTCATTTTTGAATTTGAGTGG	scaffold_2: 6571746-6571769
	nBDEL_LG_R	CCACCACAAAGAATAACGATACC	scaffold_2: 12563720- 12563698
Deletion 2-2	nBDEL_SM_F	CGGTACAATAAGTTATCACTATTCAGAG	scaffold_2: 12563142-12563169
	nBDEL_SM_R	CAGCAACGTTATCGCATGTAAA	scaffold_2: 6605253-6605232
Inversion 4-1	nBINV_LFT_F	CCCACCAAGTTCATGATTCCTTA	scaffold_4: 3540164-3540186
	nBINV_LFT_R	AGTGCTTCACTTTGCGAGAG	scaffold_4: 7515130-7515149
Inversion 4-2	nBINV_RGT_F	AGGTGGAGAAGACGATGAGTA	scaffold_4: 3540590-3540570
	nBINV_RGT_R	CCCAATAAGAGGCGTAGCATAG	scaffold_4: 7515659-7515638 *

* The closest one to nBINV_RGT_F among all hits

Table 3.3 Summary of Retained Potential Chromosome Rearrangements from novoBreak

Structural variation type *	Name	Chromosome A	Breakpoint A	Chromosome B	Breakpoint B	Structural variant length
Large deletion	DEL2-6572073:12563420	scaffold_2	6572073	scaffold_2	12563420	5991347
Large deletion	DEL2-14779349:19296626	scaffold_2	14779349	scaffold_2	19296626	-4517277
Translocation	TRA1/3-6414529:36922878	scaffold_1	6414529	scaffold_3	36922878	NA
Translocation	TRA1/3-27525574:31452544	scaffold_1	27525574	scaffold_3	31452544	NA
Translocation	TRA6/9-2254577:3795908	scaffold_6	2254577	scaffold_9	3795908	NA
Translocation	TRA6/5-15831560:23570755	scaffold_6	15831560	scaffold_5	23570755	NA
Translocation	TRA9/6-3795966:2254577	scaffold_9	3795966	scaffold_6	2254577	NA
Large inversion	INV4-3540415:7515397	scaffold_4	3540415	scaffold_4	7515397	3974981
Large inversion	INV4-3540417:7515406	scaffold_4	3540417	scaffold_4	7515406	3974988
Large inversion	INV5-16675392:26058735	scaffold_5	16675392	scaffold_5	26058735	9383342

* Structural variation types are categorized directly by the prediction results of novoBreak according to its own criteria but may not represent the actual type of rearrangement.

Table 3.4 Summary of Potential Chromosome Rearrangements for Real-Time PCR Verification after Investigating Reads on IGV

Structural variation type *	Name [z]	Chromosome A	Breakpoint A	Chromosome B	Breakpoint B	Structural variation length
Large deletion	DEL2	scaffold_2	6572073	scaffold_2	12563420	5991347
Translocation	TRA1/3	scaffold_1	27525574	scaffold_3	31452544	NA
Large inversion	INV4	scaffold_4	3540415	scaffold_4	7515397	3974981

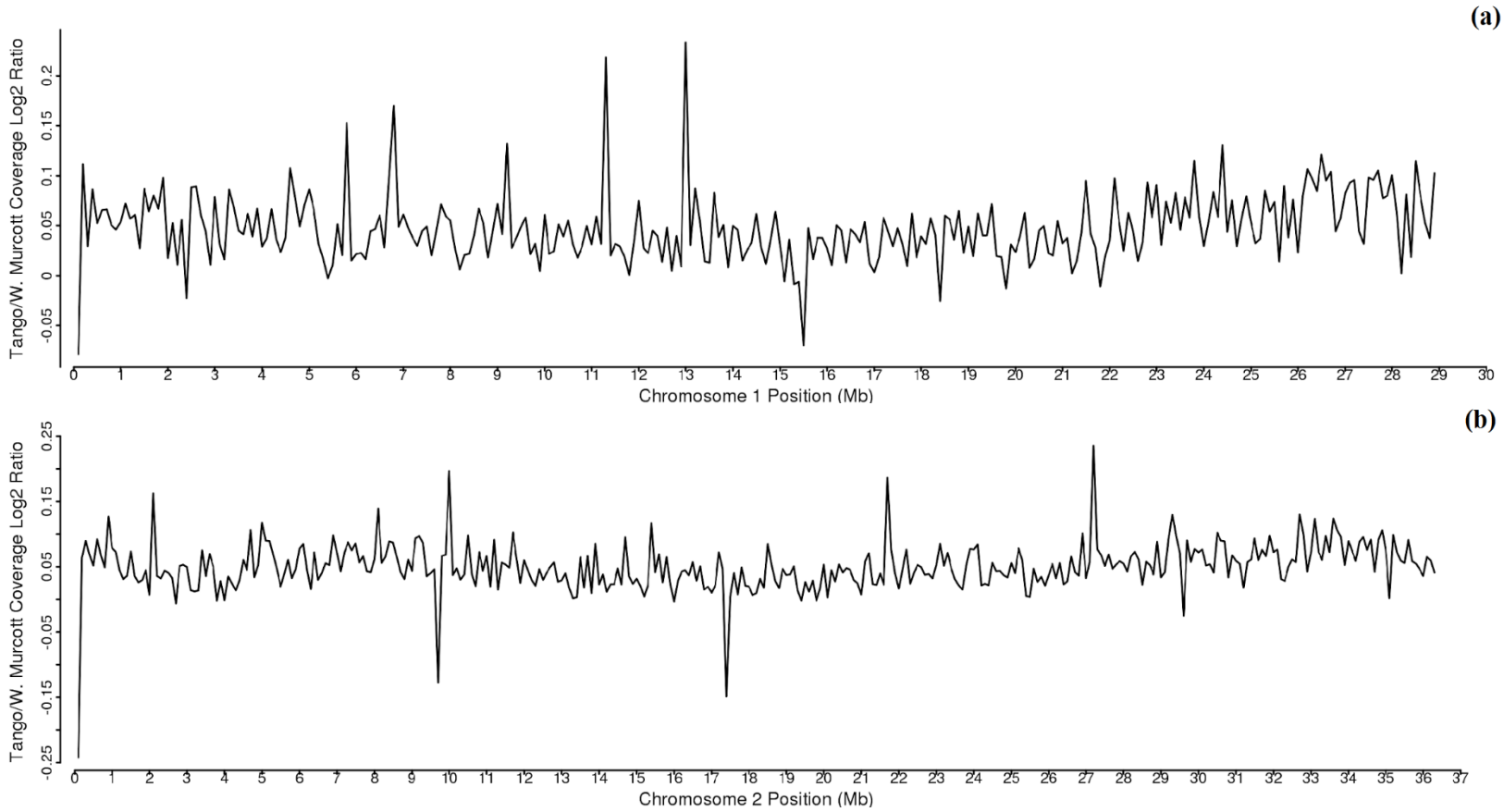
* Structural variation types are categorized directly by the prediction results of novoBreak according to its own criteria but may not represent the actual type of rearrangement.

[z] DEL2 is the alternative name for DEL2-6572073:12563420, TRA1/3 is the alternative name for TRA1/3-27525574:31452544, and INV4 is the alternative name for the combination of INV4-3540415:7515397 and INV4-3540417:7515406.

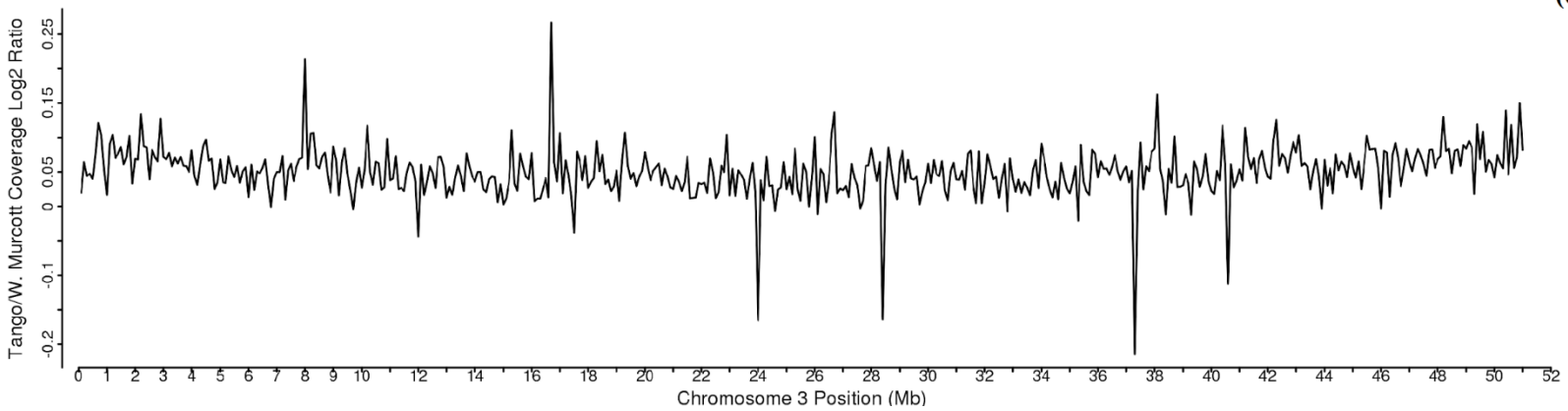
Table 3.5 Summary of Potential Genes Affected by Breakpoints of Potential Chromosome Rearrangements

Structural variation type *	Name	Breakpoint location	Predicted gene that includes the breakpoint
Large deletion	DEL2	scaffold_2: 6572073	Not in a gene
		scaffold_2: 12563420	Ciclev10014550m.g: acyltransferase-like protein At3g26840
Translocation	TRA1/ 3	scaffold_1: 27525574	Not in a gene
		scaffold_3: 31452544	Not in a gene
Large inversion	INV4	scaffold_4: 3540415	Ciclev10030974m.g: leukotriene A-4 hydrolase homolog
		scaffold_4: 7515397	Not in a gene

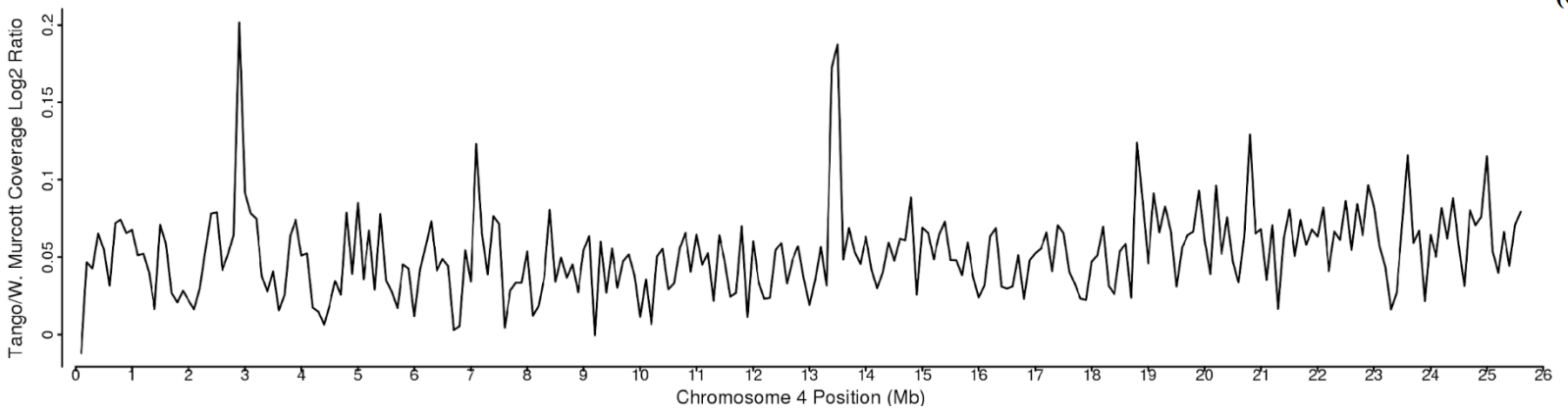
* Structural variation types are categorized directly by the prediction results of novoBreak according to its own criteria but may not represent the actual type of rearrangement.


(a)
Tango/W. Murcott Coverage Log2 Ratio
-0.05
0
0.05
0.1
0.15
0.2
0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30
Chromosome 1 Position (Mb)
(b)
Tango/W. Murcott Coverage Log2 Ratio
-0.25
-0.15
-0.05
0.05
0.15
0.25
0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37
Chromosome 2 Position (Mb)

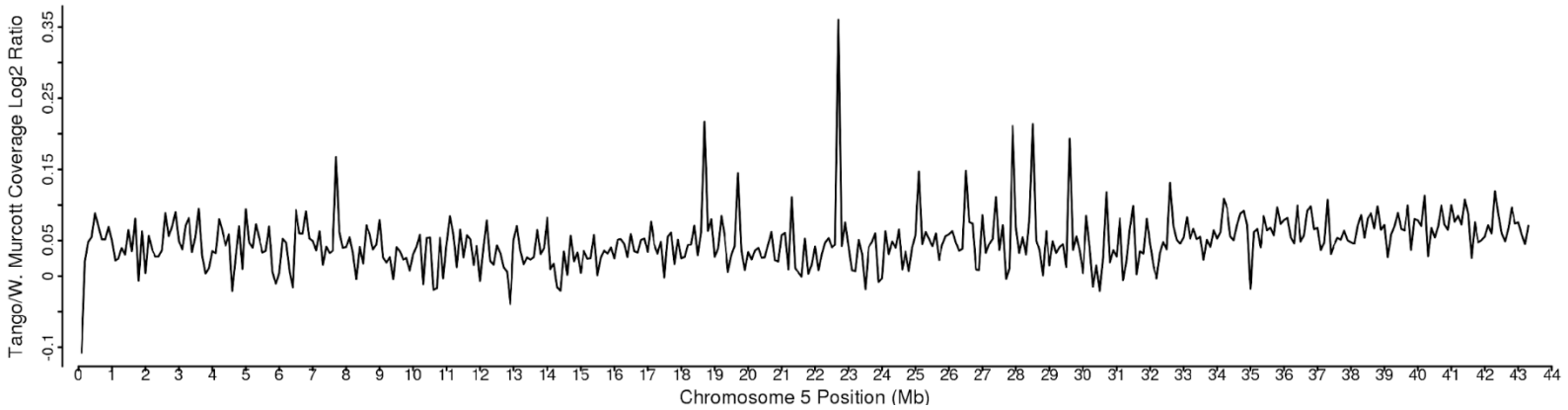
(c)


Tango/W. Murcott Coverage Log2 Ratio
-0.2
-0.1
0
0.05
0.15
0.25
0 1 2 3 4 5 6 7 8 9 10 12 14 16 18 20 22 24 26 28 30 32 34 36 38 40 42 44 46 48 50 52
Chromosome 3 Position (Mb)

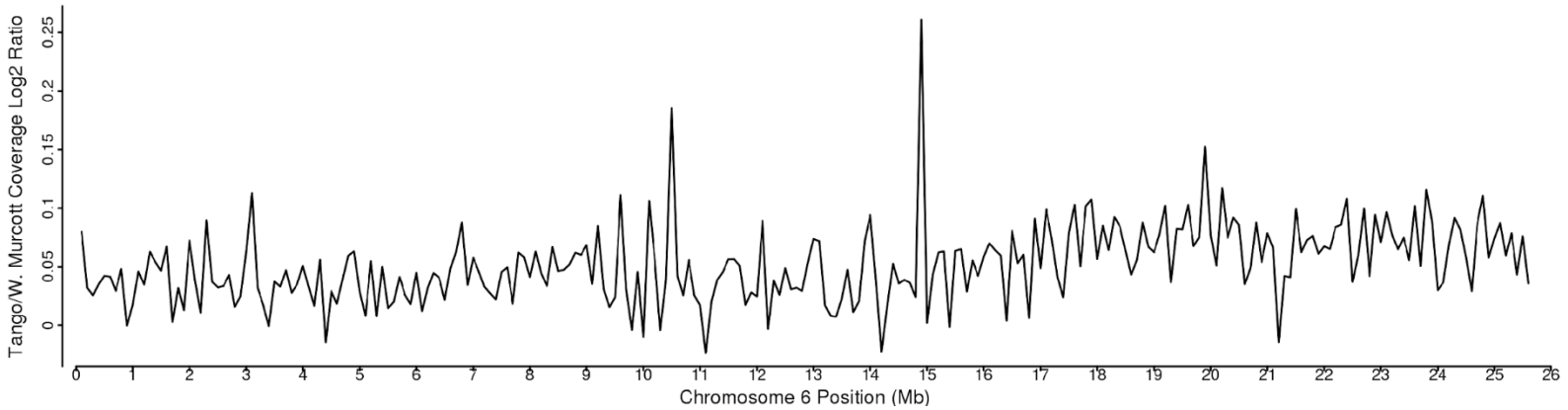
(d)


Tango/W. Murcott Coverage Log2 Ratio
0
0.05
0.1
0.15
0.2
0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26
Chromosome 4 Position (Mb)

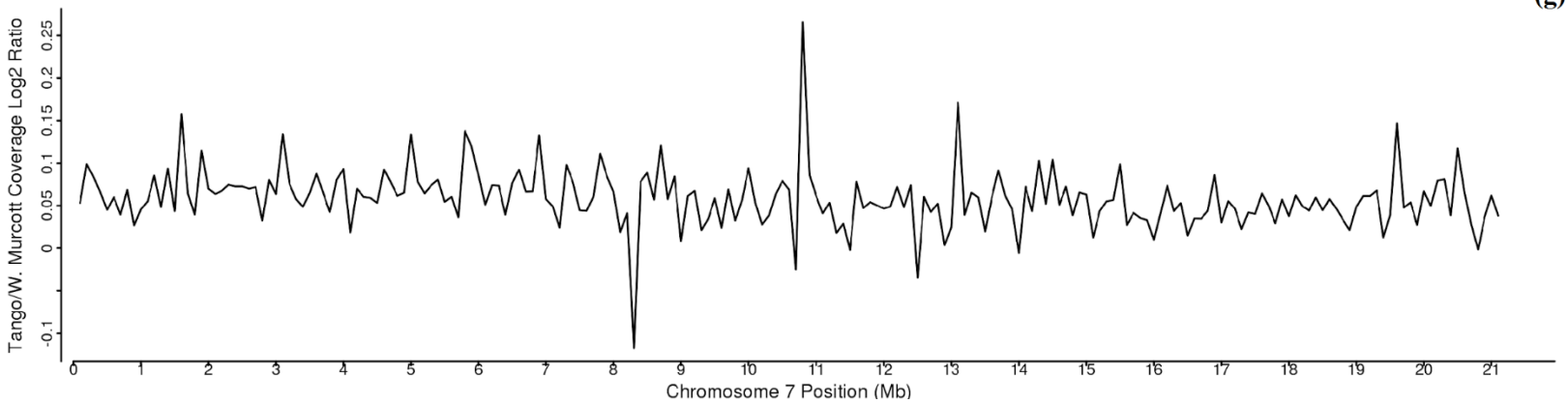
(e)


Tango/W. Murcott Coverage Log2 Ratio
-0.1
0
0.05
0.15
0.25
0.35
0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44
Chromosome 5 Position (Mb)

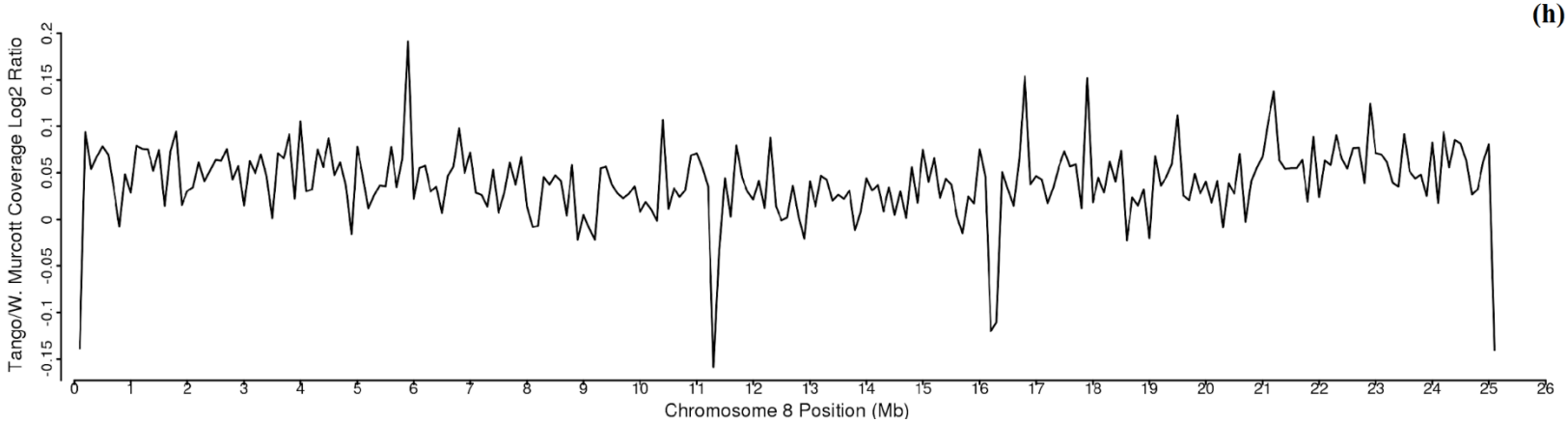
(f)


Tango/W. Murcott Coverage Log2 Ratio
0
0.05
0.1
0.15
0.2
0.25
0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26
Chromosome 6 Position (Mb)

(g)


Tango/W. Murcott Coverage Log2 Ratio
-0.1
0
0.05
0.1
0.15
0.2
0.25
0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21
Chromosome 7 Position (Mb)

(h)


Tango/W. Murcott Coverage Log2 Ratio
-0.15
-0.1
-0.05
0
0.05
0.1
0.15
0.2
0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26
Chromosome 8 Position (Mb)

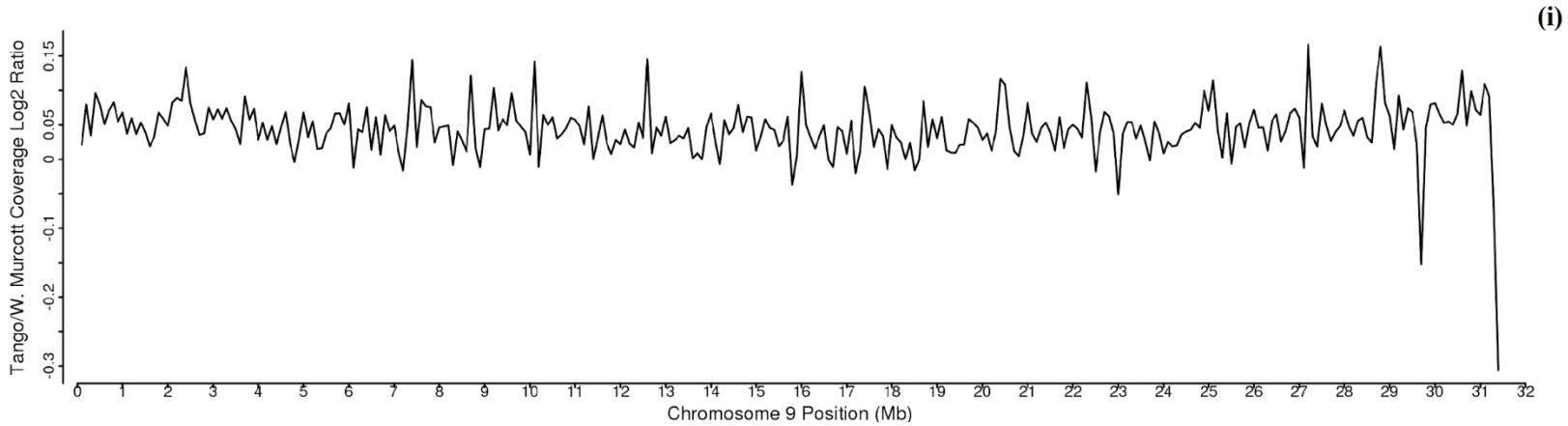


Figure 3.1 Log2 Ratio of Read Depth of Tango versus W. Murcott. (a) – (i) Nine chromosomes are divided into consecutive windows of 100 kb in size. The alignment read depth of Tango and W. Murcott within each window were counted and the log2 ratio was calculated as $log_2 \frac{Tango\ read\ depth}{W.Murcott\ read\ depth}$ and plotted on nine chromosomes.

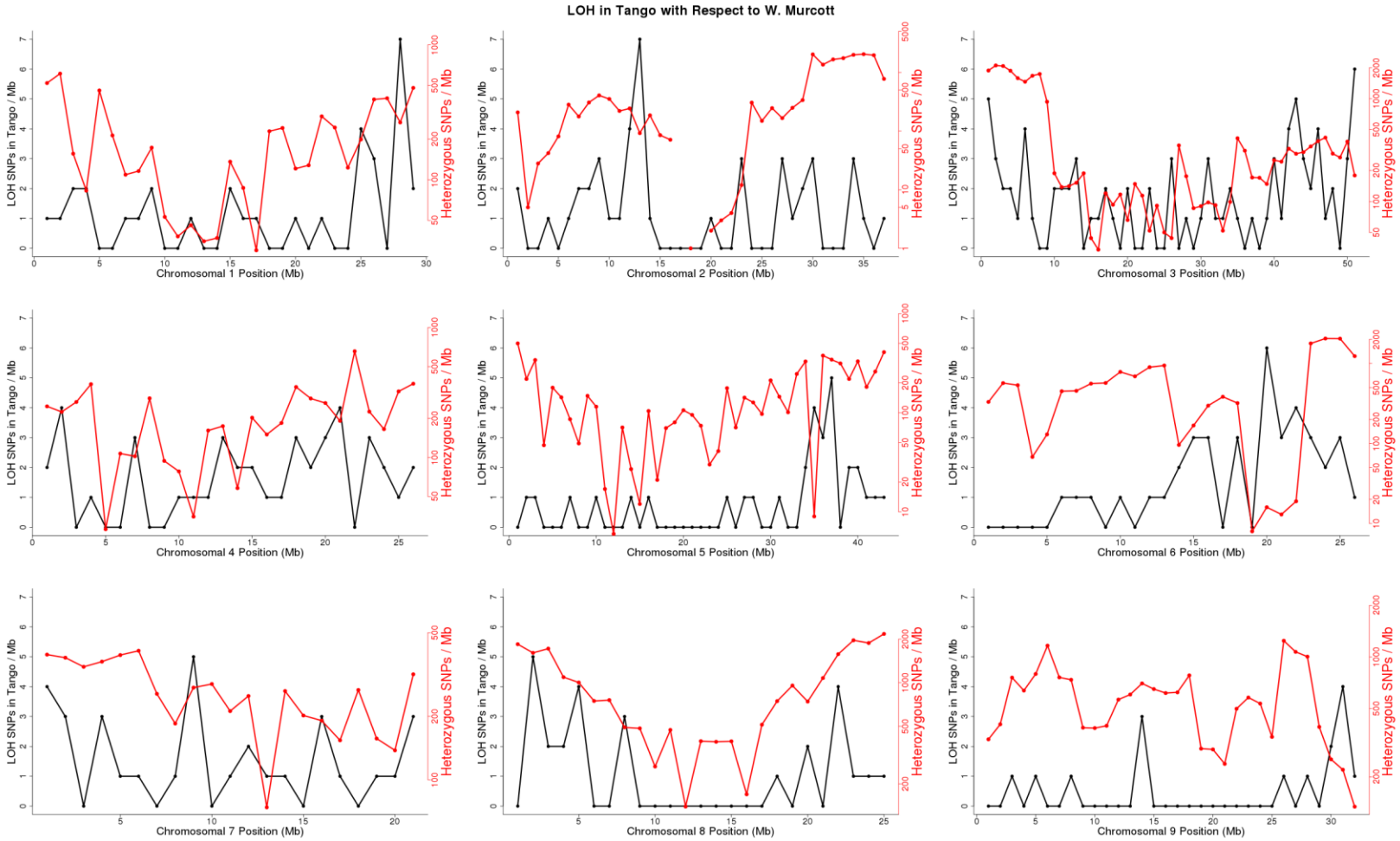


Figure 3.2 Comparison of Tango LOH Position Counts and W. Murcott High-Resolution Nuclear Genome Probes Counts from Citrus 15AX Array. Each chromosome is divided into consecutive windows of 1 Mb in size. LOH position counts in Tango (black) with respect to W. Murcott and counts of high-resolution nuclear genome probes called as heterozygous in W. Murcott (red) from *Axiom® Citrus 15AX* array within each window were compared and plotted on nine chromosomes.

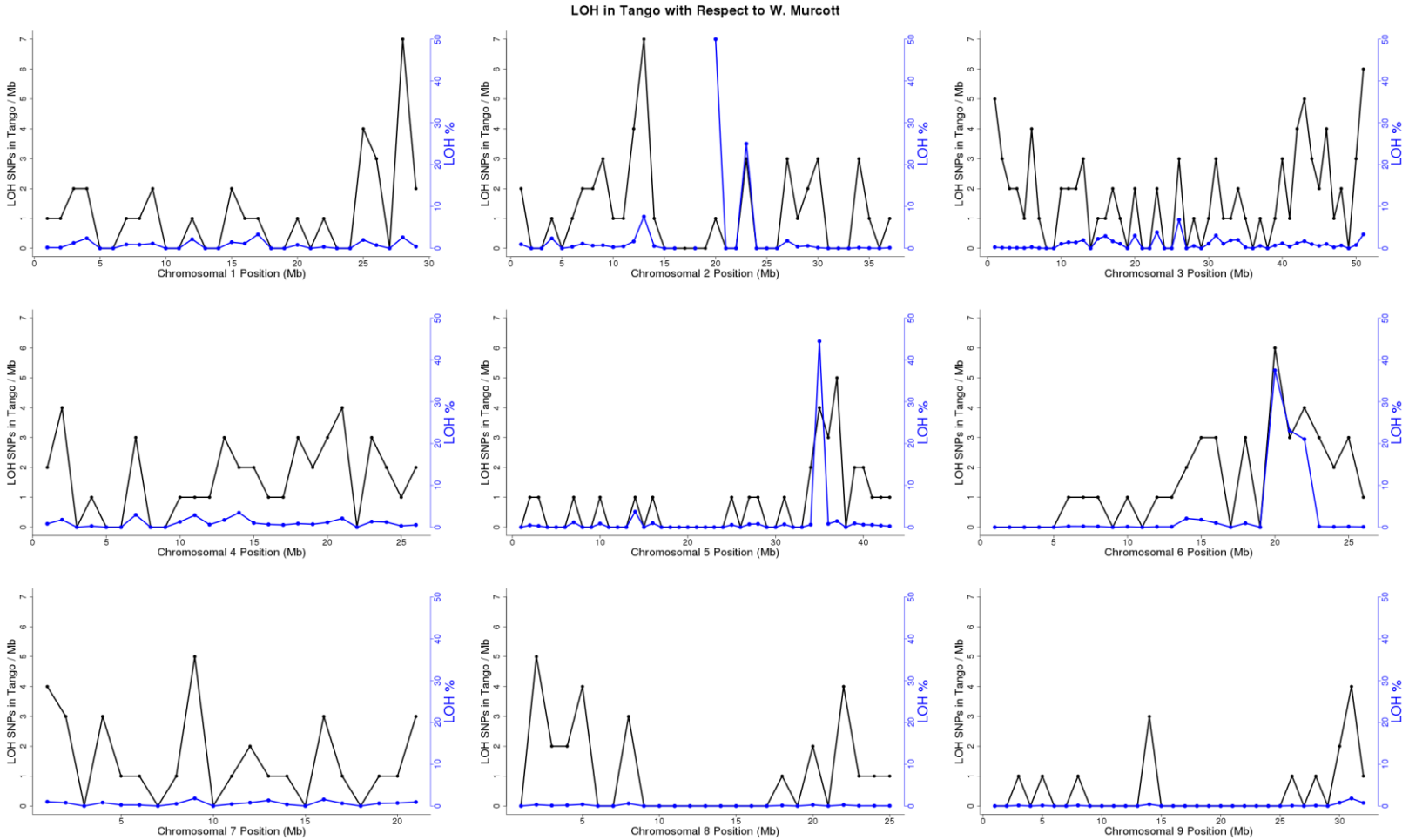
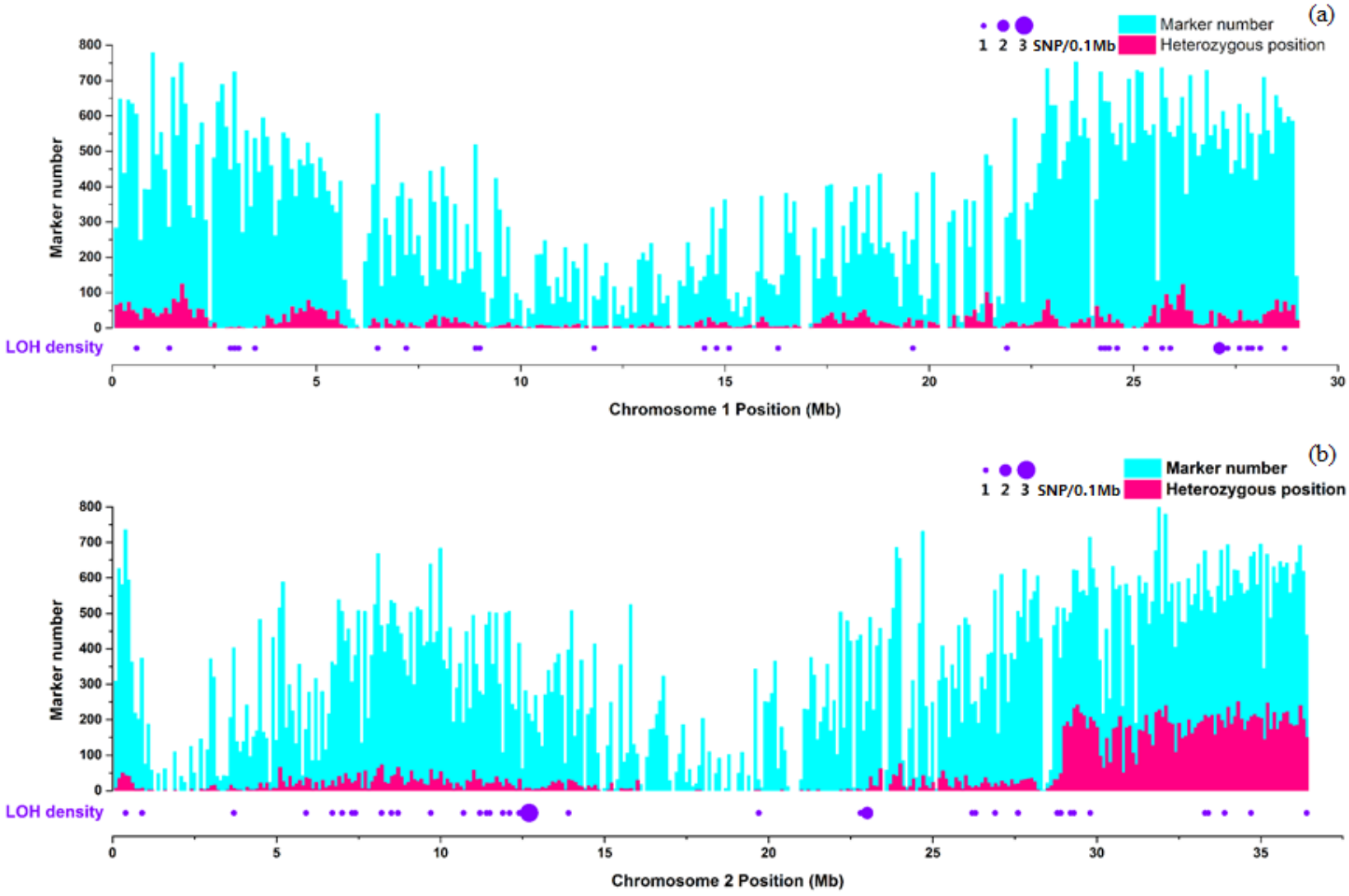
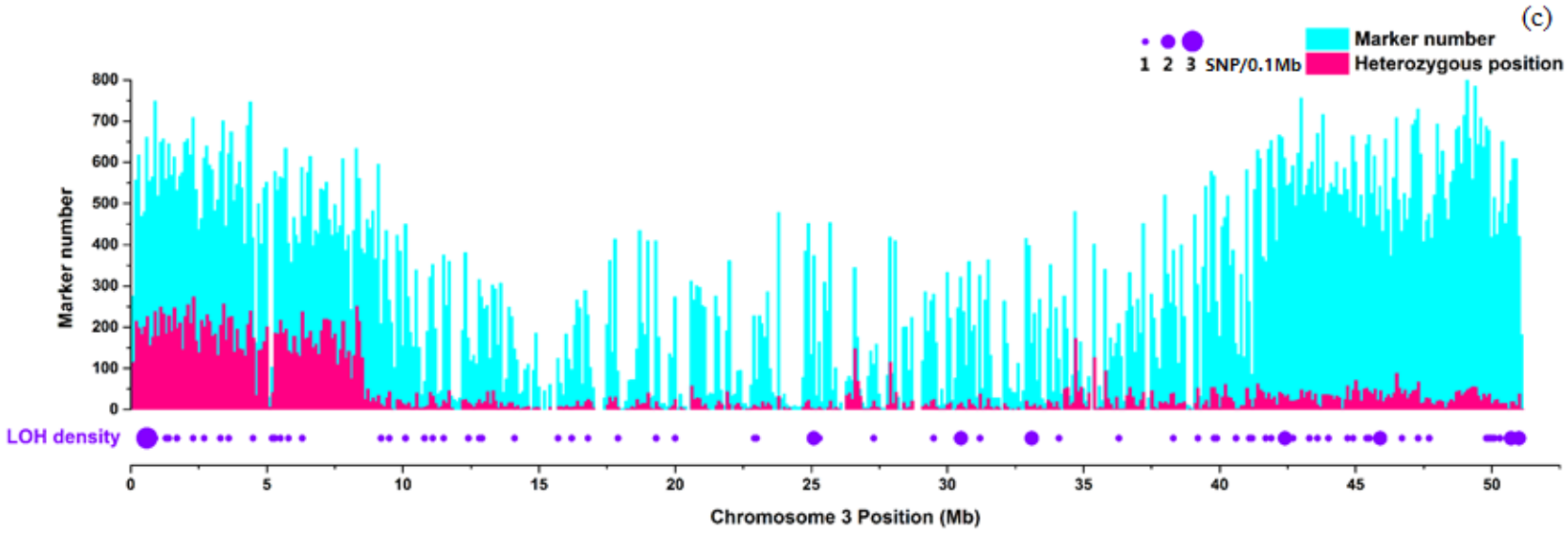
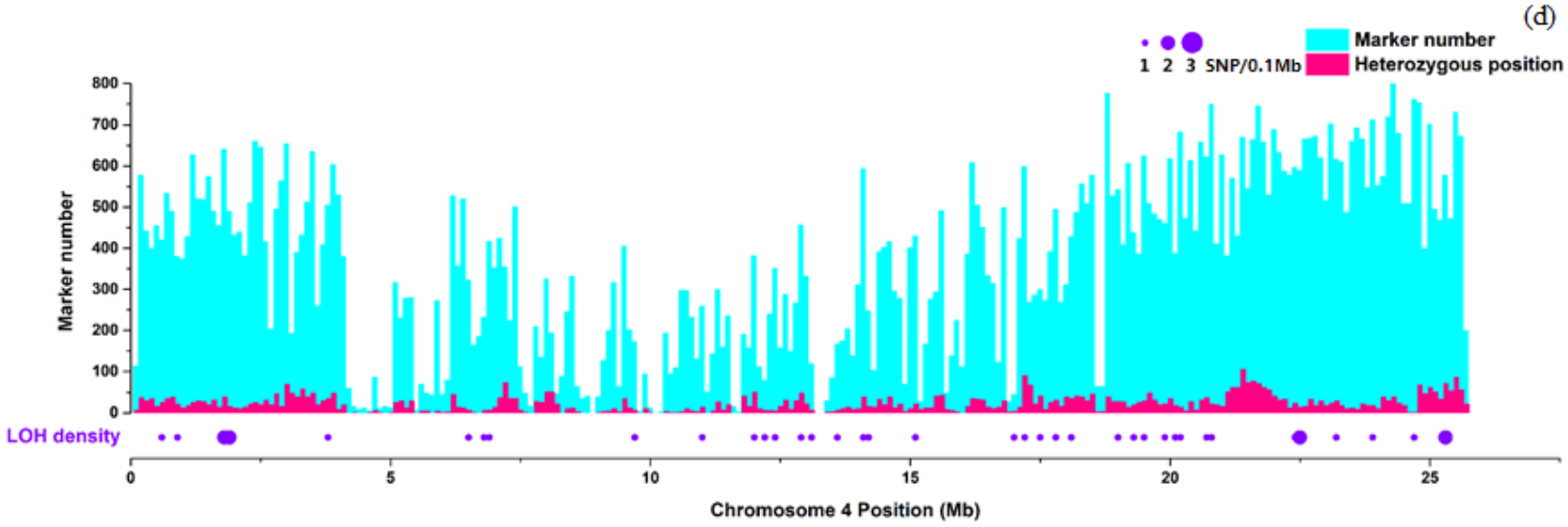
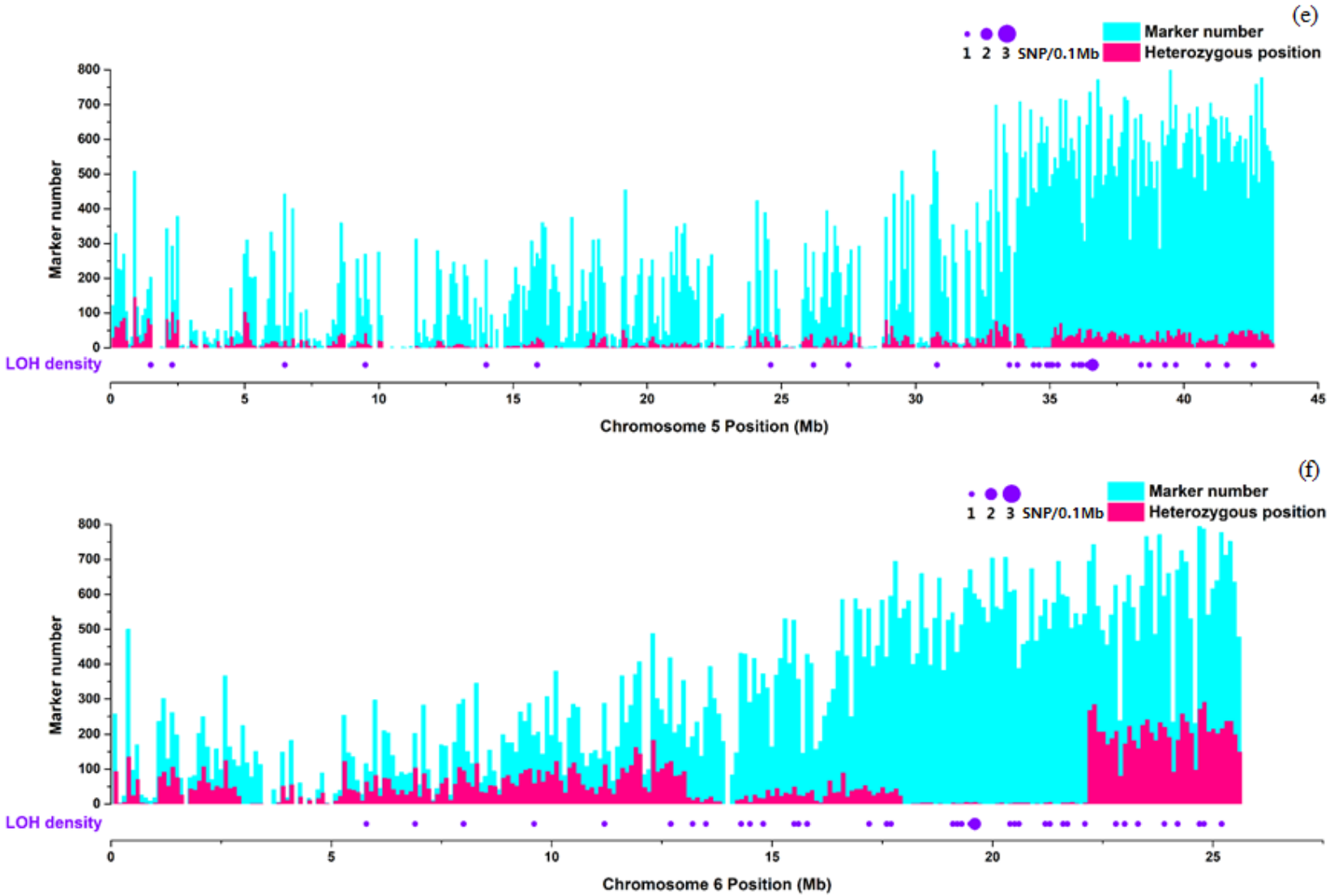


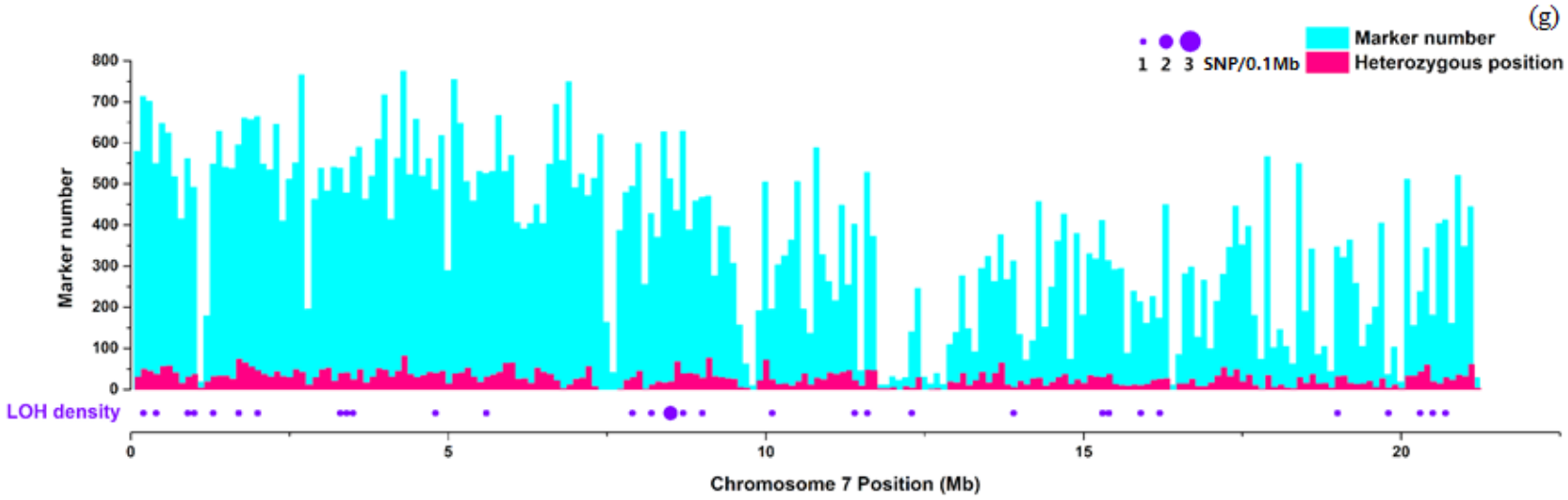
Figure 3.3 Comparison of Tango LOH Position Counts and Percentage of LOH Positions from Citrus 15AX Array. Each chromosome is divided into consecutive windows of 1 Mb in size. LOH position count in Tango with respect to W. Murcott (black) and the percentage of LOH positions of all high-resolution nuclear genome probes exhibiting heterozygous in W. Murcott (blue) from *Axiom® Citrus 15AX* array within each window were compared and plotted on nine chromosomes.

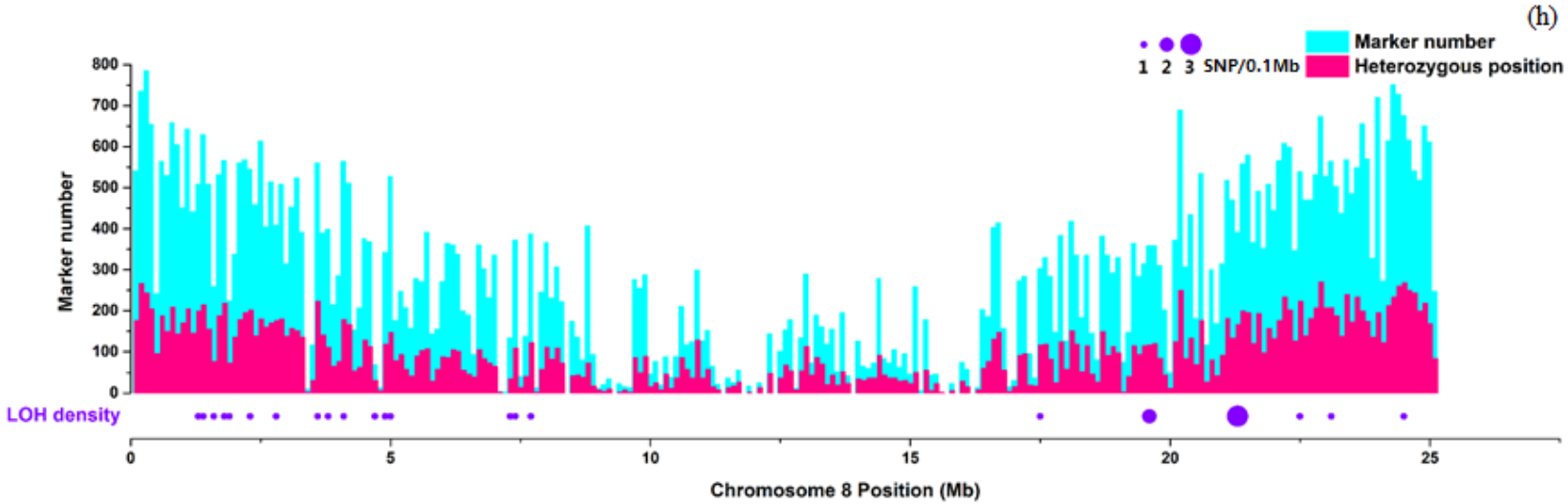

(a)
1 2 3 SNP/0.1Mb
Marker number
Heterozygous position
Marker number
LOH density
Chromosome 1 Position (Mb)
(b)
1 2 3 SNP/0.1Mb
Marker number
Heterozygous position
Marker number
LOH density
Chromosome 2 Position (Mb)


(c)
1 2 3 SNP/0.1Mb
Marker number
Heterozygous position
Marker number
LOH density
Chromosome 3 Position (Mb)


(d)
1 2 3 SNP/0.1Mb
Marker number
Heterozygous position
Marker number
LOH density
Chromosome 4 Position (Mb)


(e)
1 2 3 SNP/0.1Mb
Marker number
Heterozygous position
Marker number
LOH density
Chromosome 5 Position (Mb)
(f)
1 2 3 SNP/0.1Mb
Marker number
Heterozygous position
Marker number
LOH density
Chromosome 6 Position (Mb)


(g)
1 2 3 SNP/0.1Mb
Marker number
Heterozygous position
Marker number
LOH density
Chromosome 7 Position (Mb)


(h)
1 2 3 SNP/0.1Mb
Marker number
Heterozygous position
Marker number
LOH density
Chromosome 8 Position (Mb)

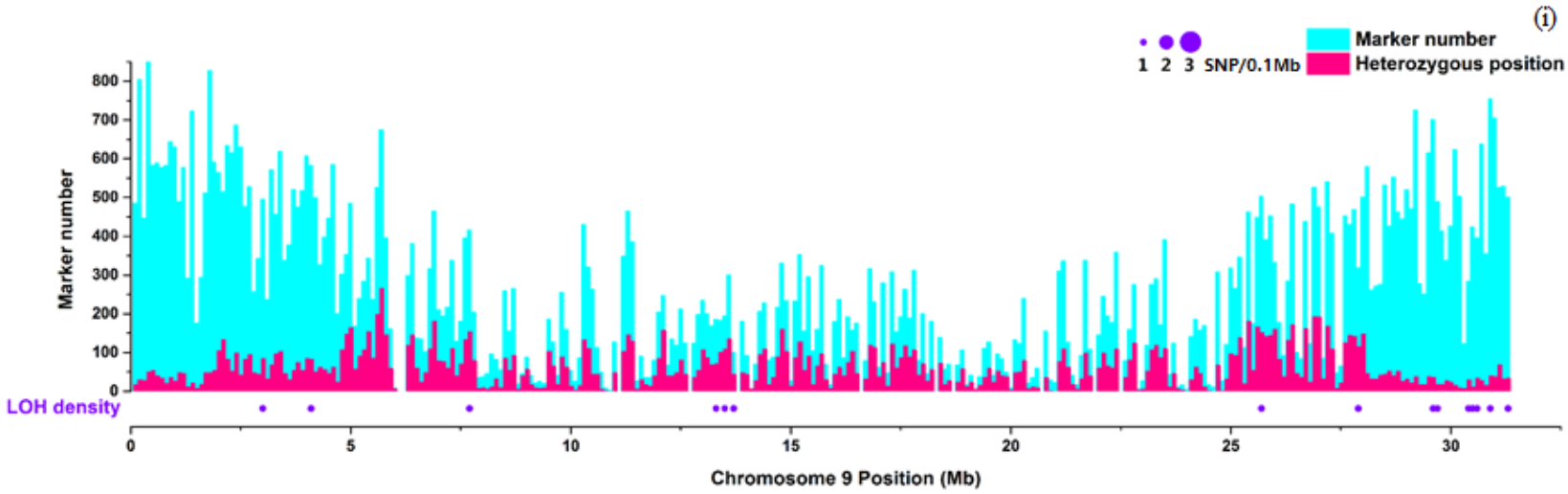


Figure 3.4 The Density of Valid SNP, Heterozygous Markers in W. Murcott and LOH Position in Tango. (a) – (i) Nine chromosomes are divided into consecutive windows of 100 kb in size. The high-resolution nuclear genome probe (SNP markers) counts (cyan), counts of high-resolution nuclear genome probes called as heterozygous in W. Murcott (pink) and LOH position counts (purple dots under histogram, in form of density, i.e., number of LOH positions per 100 kb) in Tango from *Axiom*® *Citrus 15AX* array within each window were plotted on nine chromosomes.

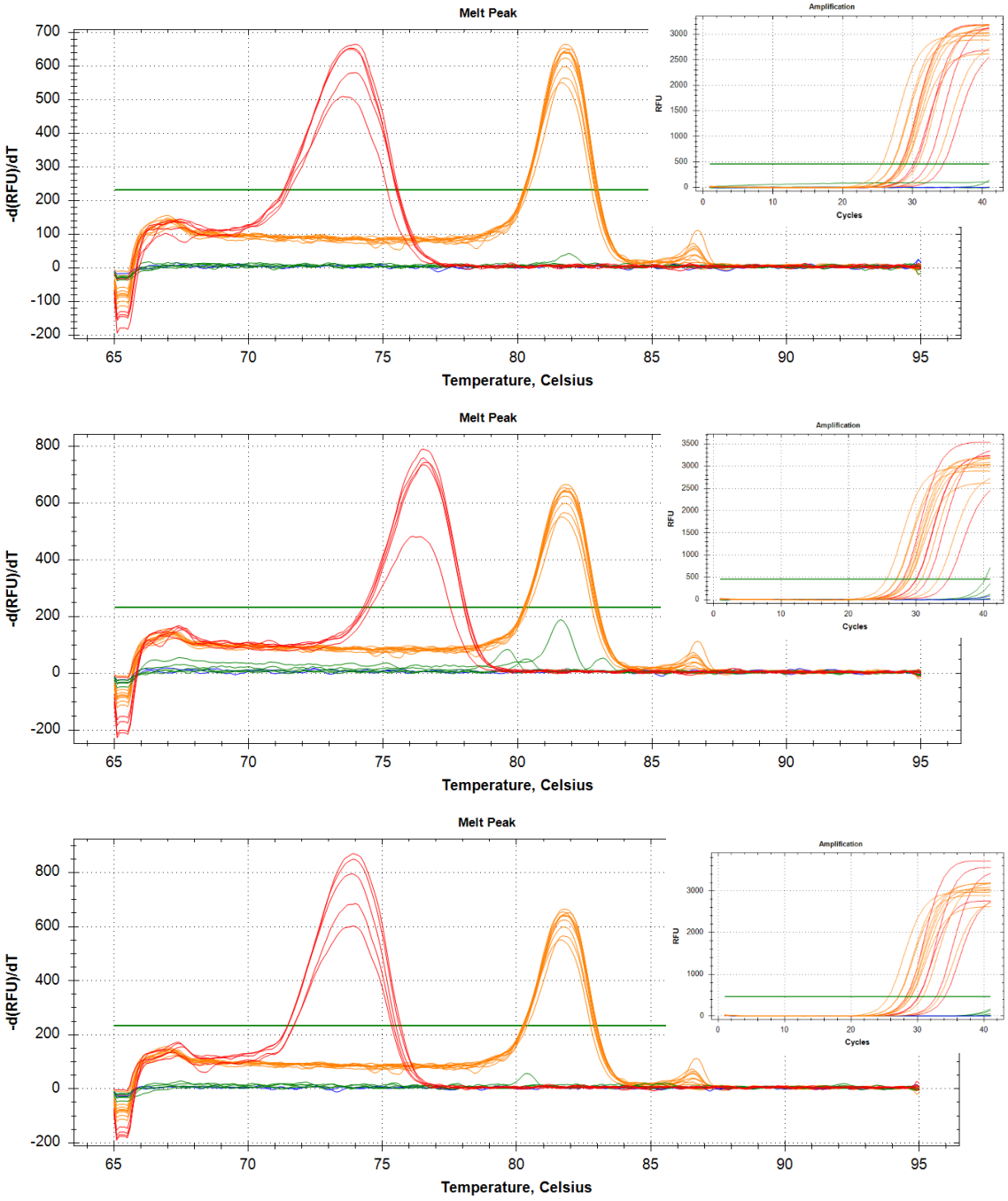


Figure 3.5 Real-Time PCR Validation for Large Deletion DEL2. The comparison of real-time PCR of five Tango leaf DNA samples with large deletion primers (red), five W. Murcott leaf DNA samples with large deletion primers (green), Tango leaf DNA with MDH primers and W. Murcott leaf DNA with MDH primers (orange) and NTC with large deletion primers (blue). Amplification curves are showed on upper right corner. Upper panel: primer pair NB81 and NB82. Middle panel: primer pair NB87 and NB88. Lower panel: primer pair NB93 and NB94.

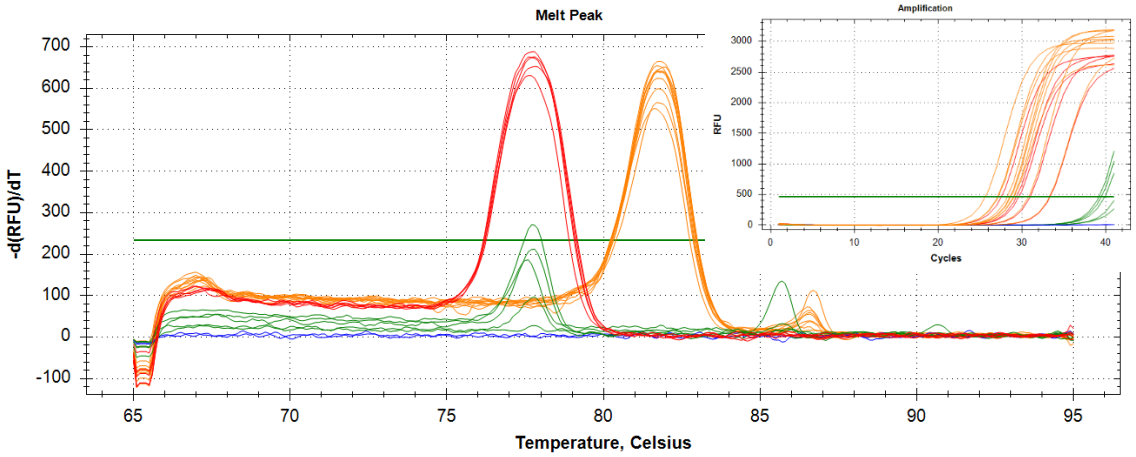


Figure 3.6 Real-Time PCR Validation for Translocation TRA1/3. The comparison of real-time PCR of five Tango leaf DNA samples with translocation primers (red), five W. Murcott leaf DNA samples with translocation primers (green), Tango leaf DNA with MDH primers and W. Murcott leaf DNA with MDH primers (orange) and NTC with translocation primers (blue). Amplification curves are showed on upper right corner. Primer pair: NB13 and NB14.

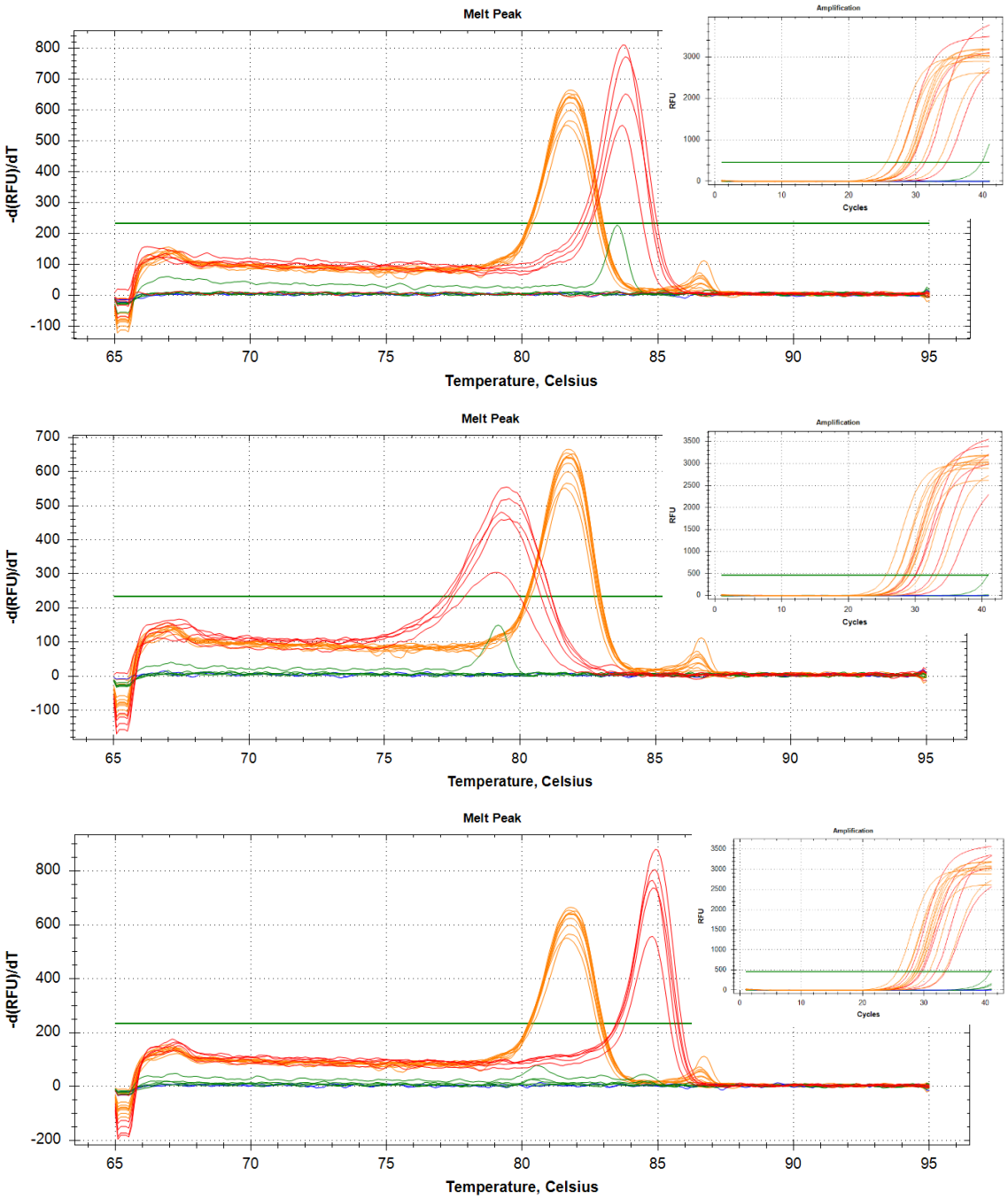


Figure 3.7 Real-Time PCR Validation for Large Inversion INV4. The comparison of real-time PCR of five Tango leaf DNA samples with large inversion primers (red), five W. Murcott leaf DNA samples with large inversion primers (green), Tango leaf DNA with MDH primers and W. Murcott leaf DNA with MDH primers (orange) and NTC with large inversion primers (blue). Amplification curves are showed on upper right corner. Upper panel: primer pair NB143 and NB144. Middle panel: primer pair NB115 and NB116. Lower panel: primer pair NB153 and NB154.

```
 Score =  523 bits (264), Expect = e-147
 Identities = 315/330 (95%), Gaps = 9/330 (2%)
 Strand = Plus / Plus

Query: 1       ttcttcatttttgaatttgagtggccaacaaataaaattatttgaaaactactctatcct 60
                |||||||||||||||||||||||||||||||||||||||||||||||||| |||||||||
Sbjct: 6571746 ttcttcatttttgaatttgagtggccaacaaataaaattatttgaaaactgctctatcct 6571805


Query: 61      gagagttattggtagatgtgtgattgctaa-------ccaccagaaaaaa-taacaaaaa 112
                |||||||||||||||||  ||||||||||       |||| |||||||| |||||||||
Sbjct: 6571806 gagagttattggtagatgcatgattgctaaattctaaccacaagaaaaaaataacaaaaa 6571865


Query: 113     agaaatttaattgaggaaaatcaaatatctataaagatttgtaccaatcatgtacacaaa 172
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct: 6571866 agaaatttaattgaggaaaatcaaatatctataaagatttgtaccaatcatgtacacaaa 6571925


Query: 173     tcaaactccctcagtgttctaaagtgtaaaggccagagaagaccttgactcactattgtg 232
                |||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||
Sbjct: 6571926 tcaaactccctcagtgttctaaagtgtaaagaccagagaagaccttgactcactattgtg 6571985


Query: 233     tttgaaaacgagtgttgcta-ttcaacgcgggaagaagatgattatgggcctgctcatgg 291
                ||||||||| |||||||||| ||||||||||||||||||||||||||||||||||||||
Sbjct: 6571986 tttgaaaacaagtgttgctatttcaacgcgggaagaagatgattatgggcctgctcatgg 6572045


Query: 292     tgtccttttcctgggttaaagtgtctatat 321
                |||||||||||||||||||||||||||||
Sbjct: 6572046 tgtccttttcctgggttaaagtgtctatat 6572075
Score =  595 bits (300), Expect = e-168
Identities = 300/300 (100%)
Strand = Plus / Plus


Query: 319     tatttaaattttgaagggattgatggagtaggacttggactggttaggcaccattactc 378
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct: 12563421 tatttaaattttgaagggattgatggagtaggacttggactggttaggcaccattactc 12563480


Query: 379     gcttggaaagtaagtaatgtccaagatttgagtgctttataccatcaatcattgataatc 438
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct: 12563481 gcttggaaagtaagtaatgtccaagatttgagtgctttataccatcaatcattgataatc 12563540


Query: 439     agaactgttgagaagttttcgaaattgaagccatcttgatgtggacatccaatgattcac 498
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct: 12563541 agaactgttgagaagttttcgaaattgaagccatcttgatgtggacatccaatgattcac 12563600


Query: 499     catgtatgtctgagtaatttatcgatttttttattaatgaaggatttttgatatatggtg 558
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct: 12563601 catgtatgtctgagtaatttatcgatttttttattaatgaaggatttttgatatatggtg 12563660


Query: 559     cttgcatattcctgtcaaggatcgaacgtcgtttgcaggtatcgttattctttgtggtgg 618
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct: 12563661 cttgcatattcctgtcaaggatcgaacgtcgtttgcaggtatcgttattctttgtggtgg 12563720
```

Figure 3.8 BLAST Result of Deletion 2 - Breakpoint 1. 1st bp to 321st bp were aligned to scaffold_2 from 6571746th bp to 6572075th bp, and 319th bp to 618th bp were aligned to scaffold_2 from 12563421st bp to 12563720th bp. The TAT trinucleotide from 319th nucleotide to 321st nucleotide at the breakpoint is shown in red box.

```
 Score =  319 bits (161), Expect = 6e-86
 Identities = 222/238 (93%), Gaps = 7/238 (2%)
 Strand = Plus / Plus

Query: 16       aaggcatttgtattatgta-tgtttgattcataaag-gcattgcacatatga-atggtaa 72
                 |||||||||||||||||| |||||||||||||||  ||||||||||||||| |||||||
Sbjct: 12563183 aaggcatttgtattatgtaatgtttgattcataaaacgcattgcacatatgaaatggtaa 12563242


Query: 73       tggactcct----tggtttataatgctgtagtatgttatatatgttaagtcattatagct 128
                   || |  |    |||||||||||||||||||||||| |||||||  ||||||||||||
Sbjct: 12563243 aagattaatagattggtttataatgctgtagtatgttagatatgttttgtcattatagct 12563302


Query: 129      caattgttgatttgccaatttagaatttatttgttgctggagtattttctttttttcaaat 188
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct: 12563303 caattgttgatttgccaatttagaatttatttgttgctggagtattttctttttttcaaat 12563362


Query: 189      ggtttacttatagctttgaagctctgagtggaaatagcatgaattattcatttttttt 246
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct: 12563363 ggtttacttatagctttgaagctctgagtggaaatagcatgaattattcatttttttt 12563420
 Score =  196 bits (99), Expect = 6e-49
 Identities = 120/127 (94%)
 Strand = Plus / Plus


Query: 246      tatgaactttataagaacttgcagaatccaatcaaccgaaaatttacaatcatggatccg 305
                 ||||||||||||||||||||||||||||||||||||||||| ||||||||||||||||
Sbjct: 6572073 tatgaactttataagaacttgcagaatccaatcaaccgaaaaattacaatcatggatccg 6572132


Query: 306      acttccttcattttcagattggcggcccttgtatgggctgtaatatttacatgcgataac 365
                  | |||| |||||||||||||||||||||||||| || |||||||||||||| ||||
Sbjct: 6572133 gcatcctccattttcagattggcggcccttgtatgggttgcaatatttacatgcgctaac 6572192


Query: 366      gttgctg 372
                 |||||||
Sbjct: 6572193 gttgctg 6572199
```

Figure 3.9 BLAST Result of Deletion 2 - Breakpoint 2. 16th bp to 246th bp were aligned to scaffold_2 from 12563183th bp to 12563420th bp, and 246th bp to 372nd bp were aligned to scaffold_2 from 6572073rd bp to 6572199th bp. The T at 246th nucleotide at the breakpoint is shown in red box. Bases 1-15 are not aligned due to low quality of Sanger sequencing.

```
 Score =  500 bits (252), Expect = e-140
 Identities = 252/252 (100%)
 Strand = Plus / Plus


Query: 1       cccaccaagttcatgattccttaaattccaatctcctctctcaaacaatcaacccccgag 60
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct: 3540164 cccaccaagttcatgattccttaaattccaatctcctctctcaaacaatcaacccccgag 3540223


Query: 61      aaaacactcactcagtcctccctcactgaccgggttaactcaatttccaccataacaaca 120
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct: 3540224 aaaacactcactcagtcctccctcactgaccgggttaactcaatttccaccataacaaca 3540283


Query: 121     atggcaccgatcgacccgcactcgttcaccgagtcaactcacccgctgaccacacacatc 180
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct: 3540284 atggcaccgatcgacccgcactcgttcaccgagtcaactcacccgctgaccacacacatc 3540343


Query: 181     tccctctctctctacttcgacttctcttcatccaccatccacgccgccgccatccttacc 240
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct: 3540344 tccctctctctctacttcgacttctcttcatccaccatccacgccgccgccatccttacc 3540403


Query: 241     ctcgcctccca 252
                ||||||||||||
Sbjct: 3540404 ctcgcctccca 3540415
 Score =  458 bits (231), Expect = e-127
 Identities = 258/267 (96%)
 Strand = Plus / Minus


Query: 250     ccaatattggtggttatcgacaacgatctaccatagcctttgatagccctcctagattgg 309
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct: 7515397 ccaatattggtggttatcgacaacgatctaccatagcctttgatagccctcctagattgg 7515338


Query: 310     cccgctaaagtcggtccgcttatgatcattcggacatacacgcctttcttatgtccaggg 369
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct: 7515337 cccgctaaagtcggtccgcttatgatcattcggacatacacgcctttcttatgtccaggg 7515278


Query: 370     ggtacctctcgttcaggactgcaatcagctacccgcttgccataatcctgttccgtaaat 429
                |||||||||||||||||||| |||||||||||||||| | |||||||||||| ||
Sbjct: 7515277 ggtacctctcgttcaggactgccatcagctacccgcttgcctttatcctgttccgtagat 7515218


Query: 430     ttcggagtttctctgatgccagcaacaaattcttttaggtacccgctgagcataaaagat 489
                ||| |||||||||| |||||||||||||||||||| |||||||||||||| || ||||||
Sbjct: 7515217 ttctgagtttctcttatgccagcaacaaattctttgaggtacccgctgagaatcaaagat 7515158


Query: 490     tcaatttcctctcgcaaagtgaagcac 516
                |||||||||||||||||||||||||||
Sbjct: 7515157 tcaatttcctctcgcaaagtgaagcac 7515131
```

Figure 3.10 BLAST Result of Inversion 4 - Breakpoint 1. 1st bp to 252nd bp were aligned to scaffold_4 from 3540164th bp to 3540415 th bp, and 250th bp to 516th bp were aligned to the reverse complement of scaffold_4 from 7515131st bp to 7515397th bp. The CCA trinucleotide from 250th nucleotide to 252nd nucleotide at the breakpoint is shown in red box.

```
 Score =  321 bits (162), Expect = 2e-86
 Identities = 171/174 (98%)
 Strand = Plus / Minus


Query: 1       aggtggagaagacgatgagtacagaggagtgatcggagagcgtgacgatgaggtggcggc 60
                |||||||||||||||||||||||||||||||||||||||| ||||||||||||||||||
Sbjct: 3540590 aggtggagaagacgatgagtacagaggagtgatcggagagagtgacgatgaggtggcggc 3540531


Query: 61      ctttgattgggtcgtcggttggggagagagtgaaggggagtggggttagggtttgggggt 120
                |||||||||||||||||||||||||||||||||||||||| ||||||||||||||||||
Sbjct: 3540530 ctttgattgggtcgtcggttggggagagagtgaaggggagttgggttagggtttgggggt 3540471


Query: 121     cgagtacttggtggatggtgagggagcgggtgtccagagagagagggccggcgt 174
                |||||||||||||||||||||||||||||||||||||| ||||||||||||||||
Sbjct: 3540470 cgagtacttggtggatggtgagggagcgggtgtccagtgagagagggccggcgt 3540417

 Score =  285 bits (144), Expect = 9e-76
 Identities = 144/144 (100%)
 Strand = Plus / Plus


Query: 175     tcaaccttaatgaatagggtacaccgaaggtaccccaccatcccccgcctacccttttta 234
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct: 7515406 tcaaccttaatgaatagggtacaccgaaggtaccccaccatcccccgcctacccttttta 7515465


Query: 235     cagaaaaagatgcggagggcatctcgtacccccacgacgatgccttcataattacgttga 294
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct: 7515466 cagaaaaagatgcggagggcatctcgtacccccacgacgatgccttcataattacgttga 7515525


Query: 295     aggttgtcactagtaaggtagcaa 318
                |||||||||||||||||||||||
Sbjct: 7515526 aggttgtcactagtaaggtagcaa 7515549
```

Figure 3.11 BLAST Result of Inversion 4 - Breakpoint 2. 1st bp to 174th bp were aligned to the reverse complement of scaffold_4 from 3540417th bp to 3540590th bp, and 175th bp to 318th bp were aligned to scaffold_4 from 7515406th bp to 7515549th bp.

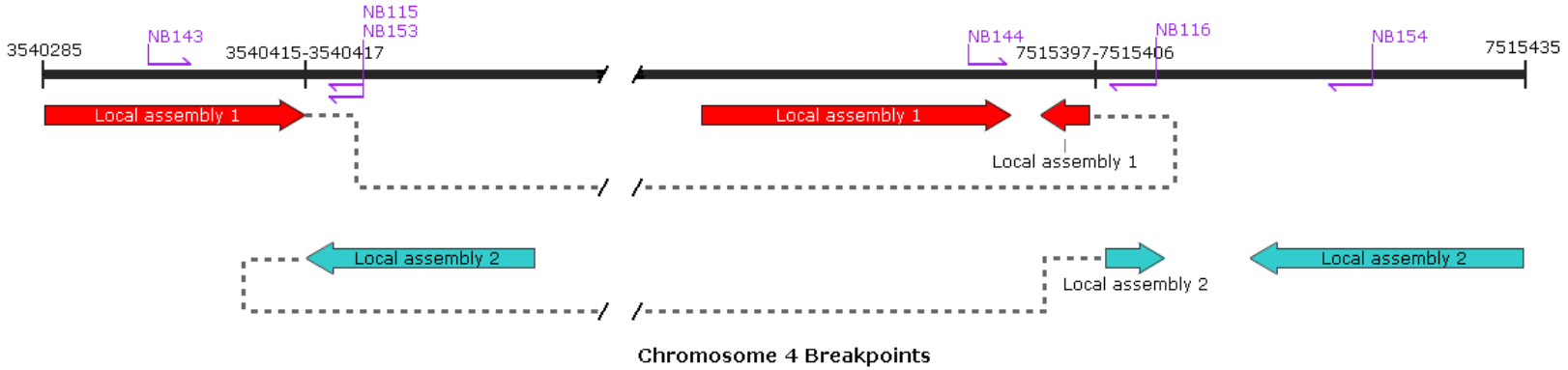


Figure 3.12 Schematic Diagram of DEL2 with Two Local Assemblies and Real-Time PCR Primer Sets. Two break points at approximately 6572073rd bp and 12563421st bp on chromosome 2 are shown on the reference genome (top black line). The approximate alignment positions of local assembly 1 (red) and local assembly 2 (teal) to the reference genome are shown, where arrows indicate there are unknown inner nucleotides between the two ends of the assembly. The dashed lines indicate the consecutive sequence of local assemblies were split-aligned to two positions on the chromosome. The position and direction of real-time PCR primer sets are shown in purple arrows.

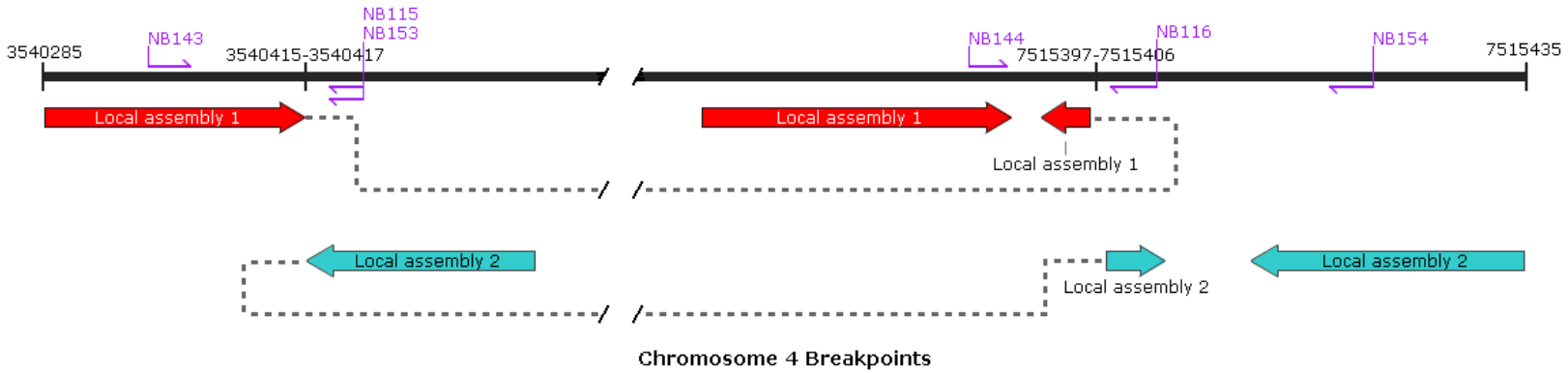


Figure 3.13 Schematic Diagram of INV4 with Two Local Assemblies and Real-Time PCR Primer Sets. Two break points at approximately 3540415th bp - 3540417th bp and 7515397th bp - 7515406th bp on chromosome 4 are shown on the reference genome (top black line). The approximate alignment positions of local assembly 1 (red) and local assembly 2 (teal) to the reference genome are shown, where arrows indicate there are unknown inner nucleotides between the two ends of the assembly. The dashed lines indicate the consecutive sequence of local assemblies were split-aligned to two positions on the chromosome. The position and direction of real-time PCR primer sets are shown in purple arrows.

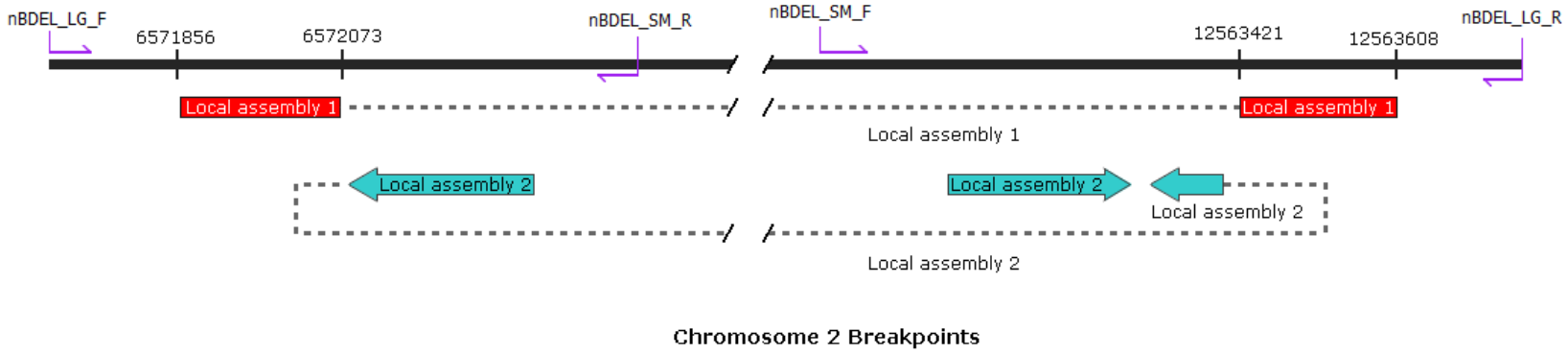


Figure 3.14 Schematic Diagram of DEL2 with Two Local Assemblies and Conventional PCR Primer Sets. Two break points at approximately 6572073rd bp and 12563421st bp on chromosome 2 are shown on the reference genome (top black line). The approximate alignment positions of local assembly 1 (red) and local assembly 2 (teal) to the reference genome are shown, where arrows indicate there are unknown inner nucleotides between the two ends of the assembly (not to scale). The dashed lines indicate the consecutive sequence of local assemblies were split-aligned to two positions on the chromosome. The position and direction of conventional PCR primer sets are shown in purple arrows (not to scale).

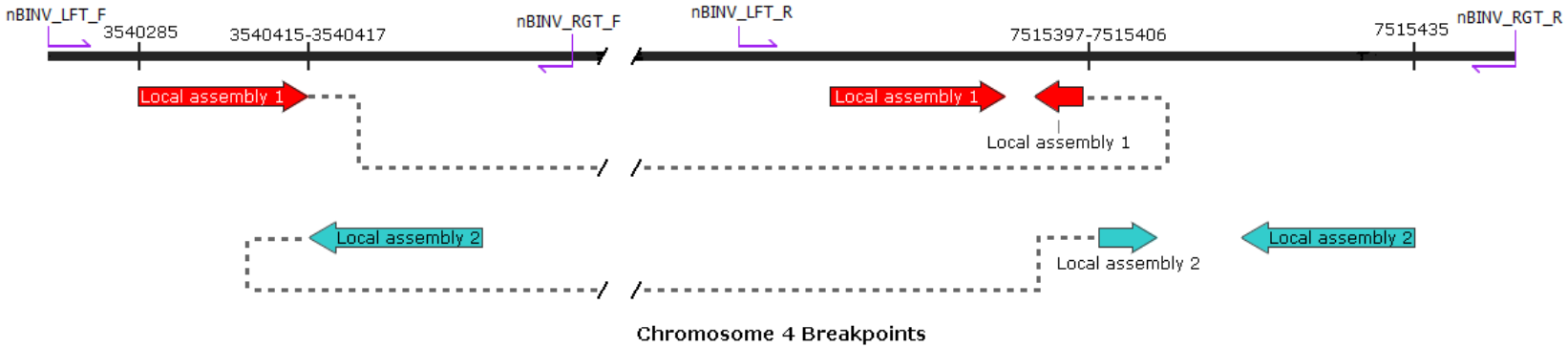


Figure 3.15 Schematic Diagram of INV4 with Two Local Assemblies and Conventional PCR Primer Sets. Two break points at approximately 3540415th bp - 3540417th bp and 7515397th bp - 7515406th bp on chromosome 4 are shown on the reference genome (top black line). The approximate alignment positions of local assembly 1 (red) and local assembly 2 (teal) to the reference genome are shown, where arrows indicate there are unknown inner nucleotides between the two ends of the assembly (not to scale). The dashed lines indicate the consecutive sequence of local assemblies were split-aligned to two positions on the chromosome. The position and direction of conventional PCR primer sets are shown in purple arrows (not to scale).

Chapter 4: Chimerism of Tango

Abstract

Tango mandarin was developed from gamma-irradiation mutation breeding of W. Murcott mandarin, and the irradiation may cause mutations in only one or two of the three cell layers of the meristem, making Tango a chimera. Because the gametes are derived from cell layer II, it is most likely that cell layer II was affected by the mutations that reduce fertility of Tango. The purpose of this study was to determine whether Tango is a chimera and which cell layers carry the molecular markers and chromosome rearrangements found previously in leaf tissue. In order to compare the genomic compositions of cell layer I and cell layer II of Tango, DNA was extracted from fruit juice vesicle and albedo tissues and tested for Indel markers based on PCR product size differences and for chromosome rearrangements using quantitative PCR. In addition, chromosome rearrangement primers were also tested on Tango pollen DNA. The results revealed chimerism in Tango, in which two short deletions, DC-1 and Del3, and three chromosome rearrangements, TRA1/3, DEL2 and INV4, were present in cell layer II derived tissues (including pollen) and absent in cell layer I of the meristem. The other short deletions (DC-7, Del2, Del4-2, Del3-3 and Del2-3) were present in both cell layer I and cell layer II of the meristem. The chromosome rearrangements present in cell layer II could be responsible for the misalignment of

chromosomes during meiosis and contribute to formation of inviable male and female gametes and very low-seeded fruit.

Introduction

When a plant has a mixture of genetically different cells growing adjacent in the plant tissue and those cells remain clearly separated, the plant is called chimera. The most common type of chimeras is variegated plants, in which cells that originated in the same shoot apical meristem have different abilities to synthesize chlorophyll, making a variegated leaf with white and green colors (PEARY et al., 1988). Graft chimera is a common chimera, in which adventitious shoots develop from a mixture of both stock and scion cells at a graft union (HANSEN et al., 2003). Most graft chimeras originate from grafting parents of the same species or genus (HANSEN et al., 2003).

Chimerism was studied and utilized long before the advent of molecular genetics (FRANK AND CHITWOOD, 2016). The first chimera recorded can be traced back to around seventeenth century when Pietro Nati discovered the growth of an adventitious shoot from the graft junction between *Citrus aurantium* and *Citrus medica* whose sectors phenotypically resembled both of the citrus progenitors, and it was eventually named 'Bizzaria' and demonstrated that it was the tissue composed of cell layers derived from two donors that lead to its unique sectored appearance (FRANK AND CHITWOOD, 2016). At the beginning of last century, Winkler performed an experiment on generating sports at the graft junction between *Solanum nigrum* and *S. lycopersicum*, in which some sports that resembled both progenitor species were found and isolated (WINKLER, 1907). The analysis of the sports indicated that their tissue was composed of separate cells from both parents

rather than from cellular fusion, and both graft-hybrids and graft-chimeras can be produced at the graft junction (WINKLER, 1907; WINKLER, 1909). Later by tracing chlorophyll inheritance in variegated geraniums, Baur proposed the model that the mature tissues within the shoot could be divided into clonally distinct parts from the shoot apical meristem (SAM), supporting Winkler's results that the phenotypic intermediates were derived from the heterogeneous arrangement of cells (FRANK AND CHITWOOD, 2016).

The shoot apical meristem is the location in which most cells of the plant body are produced. Rapid cell division in the apical meristem of an actively growing shoot makes the shoot elongate and expand (GENDRON et al., 2012). The apical meristem has a layered organization, and cell divisions are maintained in the discrete layers, leading to these layers retaining the organization into different regions during leaf and lateral bud development (BARTON AND POETHIG, 1993). The shoot apical meristem of flowering plants consists of stratified layers of clonally derived cells which is known as the "Tunica-Corpus" theory presented by Schmidt, where outer "tunica" layers divide anticlinally while an inner "corpus" layer divides both anticlinally and periclinally (ZHANG et al., 2007; ZHOU et al., 2002). There is typically a single tunica layer and a single corpus layer in gymnosperms, while in most angiosperms, there are two tunica layers and a corpus layer, named as cell layer I, cell layer II and cell layer III respectively (ZHANG et al., 2007). Cell layer I, the outmost cell layer, divides anticlinally and gives rise to the epidermis tissue, whose derivatives give rise to the colorless continuous outer covering over tissues of leaf, stem, flower petals, fruit, etc (TILNEY-BASSETT, 1986). The stigma and tissues inside the style of the ovary are also derived from cell layer I (FERNANDEZ et al., 2006). Cell layer II, which

is the cell layer under cell layer I, divides anticlinally in the dome apex and anticlinally and periclinally in the dome or shoot apical meristem flanks, the derivatives of which give rise to several layers within the stem and large proportion of the leaf blade, such as the sub-epidermal palisade mesophyll and abaxial spongy mesophyll (TILNEY-BASSETT, 1986). Cell layer III, which is in the innermost cell layer, divides in various planes, and its derivatives give rise to most of the internal tissue of the stem and cells around veins within the leaves such as deep mesophyll and vascular tissue (HANSEN et al., 2003; TILNEY-BASSETT, 1986). Cell layer III also determines the formation of a functional pedicel abscission zone and whether the meristem maintains the inflorescence state or reverts to vegetative growth (SZYMKOWIAK AND IRISH, 1999). The first demonstration of the existence of three distinct cell lineages in the shoot meristem was a study of periclinal polyploid cytochimeras (SZYMKOWIAK AND SUSSEX, 1992). Because organ formation requires the participation of all three cell layers, the coordination of proliferation is indispensable for organ primordia and shoot apical meristem integrity maintenance (HANSEN et al., 2003).

According to the 'Tunica-Corpus' theory, if a plant has genetically different cells existing in the shoot apical meristem whose cells develop into organs, this mosaic plant is called chimera (ZHANG et al., 2007). When some cells undergo mutation, either spontaneous or induced by irradiation or treatment with chemical mutagens, while the other cells remain intact, the plant becomes a chimera, specifically an autogenous chimera. If the mutated cells are located near the crest of the apical dome, all cells derived from the mutated cell by division will also carry the same mutation, resulting in genetically different cells

growing adjacent in the same plant tissue, making a chimeric plant (TIAN AND MARCOTRIGIANO, 1993). There are multiple origins of plant chimeras: 1. spontaneous or induced nuclear mutations, 2. variegated phenotypes produced by mutations in the nuclear or plastid genome that affects chloroplast development, 3. grafting, and 4. development of plantlets from callus cultures which consist of mixed cell populations (SPENA AND SALAMINI, 1995). Experimentally synthesized chimeric plants were previously used in studies of ontogenesis and cell lineages of the shoot meristem, and provided novel plant breeding approaches that combines valuable characteristics of two different cultivars (ZHOU et al., 2002). Due to the distinguishing features and values, plant chimerism has attracted considerable attention in many studies on plant development and been utilized as a tool at the cellular, tissue, and organismal level (FRANK AND CHITWOOD, 2016).

The organization of mutated cells in the apex will determine the type of chimera and the stability. Based on the spatial arrangement and genetic composition within the shoot apical meristem with heterogeneous cells, plant chimerism can be categorized into three types. Periclinal chimera is the most significant chimera, in which all cells in a single cell layer are genetically distinct (SKIRVIN AND NORTON, 2016). It occurs if the original mutated cell is located near the apical dome, and the cells produced by subsequent cell divisions form an entire layer of cells carrying the mutation, genetically different from the rest of the meristem (TIAN AND MARCOTRIGIANO, 1993). Due to the relative stability of periclinal chimeras, this type of plant can be vegetatively propagated and retain its novel characteristics (TIAN AND MARCOTRIGIANO, 1993). Mericlinal chimeras occur when the genetically different cells derived from a mutation are limited within part of a single layer

instead of covering the entire apical dome, and thus are maintained in a portion of the meristem (SKIRVIN AND NORTON, 2016). The shoots or leaves derived from that portion will carry the mutation, while shoots or leaves developed from the rest of the meristem are of the normal type (TIAN AND MARCOTRIGIANO, 1993). Usually, mericlinal chimeras are restricted to a small number of cells in a heterogenomic population of cells in one cell layer, resulting in a small portion of a leaf that may be affected (TIAN AND MARCOTRIGIANO, 1993). Sectorial chimeras result when the mutation affects a part of each layer of the apical meristem, extending from the axis to epidermis, or occurs in a non-patterned heterogenomic population of cells (FRANK AND CHITWOOD, 2016; SKIRVIN AND NORTON, 2016). Depending on the point on the apical meristem from which the shoot differentiates, a sectorial chimera may or may not generate shoots or leaves carrying the mutation (RUTH et al., 1985). Therefore, a sectorial chimera is unstable. Variegated plants can be produced by genetic mosaicism including periclinal chimerism and sectorial chimerism (MARCOTRIGIANO, 1997). For periclinal color chimerism, a critical component of a pigmentation biosynthesis pathway is mutated through a stable shoot apical layer-specific mutation, since each cell layer is responsible for a predictable part of leaf and leaf blade, and for sectorial chimerism, the pigmentation biosynthesis genes are unstably altered by highly active transposons in sectors of a shoot apical meristem (FRANK AND CHITWOOD, 2016; MARCOTRIGIANO, 1997).

Periclinal, mericlinal, and sectorial chimeras can be artificially created by several traditional and advanced transgenic techniques (FRANK AND CHITWOOD, 2016). Chimeras which result from somatic mutation are of importance in horticulture and agriculture,

because somatic mutations tend to be retained in plants with stratified meristems and plant varieties can be vegetatively propagated (SZYMKOWIAK AND SUSSEX, 1996). The pure non-chimeric mutants, which consist entirely of cells with either the L1 or L2 genotype, can be recovered by somatic embryogenesis of periclinal chimeric plants and generate new agronomically useful phenotypes (FRANKS et al., 2002). An example is that many banana cultivars were propagated from chimeras with inflorescence and leaf morphology or plant height affected by somatic mutations (WHITHAM AND SLOBODCHIKOFF, 1981).

Genetic interactions between different tissues of different genotype in chimeras have been identified in previous studies (SZYMKOWIAK AND SUSSEX, 1996). Studies showed the development of some characteristics depend solely on the genotype of their cells, such as epicarp color and juice sacs, while some other characteristics depend on interactions between the two genetically distinct tissues in the chimera, such as leaf size, stoma size, sugar content, acidity and fruit size (ZHOU et al., 2002). Periclinal chimeras are notably significant for genetic interaction studies because they disclose the interactions between cells of different layers and their derivatives, as well as the influence of epidermis and inner tissue on plant organ size and shape (ZHOU et al., 2002). Some studies showed that the exogenous epidermis of graft and periclinal chimeras imposes broad effects on many characters including distinctly different leaf size, floral morphology and fruit size and weight, and inner tissues also affect L1-derived tissue such as stomatal size and fruit acidity (ZHOU et al., 2002). This can be explained by cell displacement which occurred during leaf and fruit development when periclinal divisions of cells in the epidermal layer is followed by further anticlinal division, or delivery of plant hormones and signaling molecules

through vascular connections from genetically different cell layers and regulation of vegetative and reproductive development (ZHOU et al., 2002). In some cases, cell layer I and cell layer II may carry one or two different alleles at the same locus, resulting in the tri-allelic or tetra-allelic tissue genotypes (MONCADA et al., 2006).

Chimerism generated by mutation in the shoot apical meristem has been widely applied to viticulture. The Pinot Meunier variety was generated from a Pinot Noir shoot by spontaneous mutation where a dominant negative mutation took place within the L1 cell layers (BOSS AND THOMAS, 2002). Pixie is a counterpart variety of Pinot Meunier produced by subsequent regeneration from the L1 cell layer of Pinot Meunier, which results in a dwarfed plant and earlier flowering than Pinot Meunier (BOSS AND THOMAS, 2002). The mutational basis of horticultural sports has been extensively studied by the viticulture community. An example is that there were three microsatellite alleles at the same locus found in a Pinot Meunier, the proper interpretation of which is that a mutation occurred in one of the two alleles within some cells in the apical meristem and the mutation has been maintained in one cell layer of the apical meristem through vegetative propagation, while the original genotype is maintained in the other cell layer of the plant, and the interaction of the two genetically distinct cell layers generated the novel phenotype (FRANKS et al., 2002). In the first study of RAPD applied to detection of grape chimerism, the *in vitro* investigation showed a periclinal chimeric grape fruit had the combination of skin from cv. Pinot Noir (PN) 7613 (*Botrytis* tolerant) and pulp from cv. Chardonnay (Ch) 7535 (sensitive) (VERDISSON et al., 2015).

In citriculture, periclinal chimeras can serve as a unique type of germplasm and have been developed as a breeding method by combining a genotype that is resistant to disease and a genotype that has high fruit quality (SUGAWARA et al., 1995). In previous studies, citrus periclinal chimeras were synthesized by inarching etiolated hypocotyls from nucellar embryos of Kawano natsudaidai (*Citrus natsudaidai* Hayata) and cv. Fukuhara orange [*Citrus sinensis* (L.) Osbeck] followed by inducing calli from the horizontally cut grafted region (KUHARA, 1988; SUGAWARA et al., 1995; ZHOU et al., 2002).

The tissues of citrus fruit can be structurally divided into two main groups, the pericarp and the endocarp (OHTSU AND KUHARA, 1994). The pericarp is the fruit peel part, which can be further divided into epicarp (epidermis), hypoderm and mesocarp (OHTSU AND KUHARA, 1994). The epicarp is the outermost epidermis of the fruit peel and the hypoderm is directly beneath the epicarp with two or more layers of cells (OHTSU AND KUHARA, 1994). Under the hypoderm, there is the mesocarp, which consists of outer mesocarp (flavedo) and inner mesocarp (albedo) (OHTSU AND KUHARA, 1994). The endocarp is the pulp segments (OHTSU AND KUHARA, 1994). During fruit development, cell layer I contributes to juice sacs of fruit, and pericarp epidermis (ZHANG et al., 2007). Cell layer II contributes to a layer of small cells immediately adjacent to the epidermal cell layers, seeds, segment walls, hypoderm, and mesocarp, and influences fruit shape, navel, and color and aroma of the rind (GILLASPY et al., 1993; ZHANG et al., 2007). And cell layer III develops into the vascular bundle (ZHANG et al., 2007). Because the branches of vascular bundles permeate the mesocarp, extend to the hypoderm and terminate within the inner layers of the hypoderm, the inner hypoderm includes tissues produced from both cell layer II and

cell layer III (OHTSU AND KUHARA, 1994). The size of the floral meristem during carpel initiation and the carpel number are determined by the size of cell layer III in the shoot apical meristem (SATINA AND BLAKESLEE, 1943; SZYMKOWIAK AND SUSSEX, 1992).

There are various techniques to assess chimerism. Analysis of differences between parent and mutant genomes is the basic principle in the detection of chimerism. Due to the fast, sensitive and robust nature of PCR amplification, this technique is preferred for studies on genome identity of individuals or unique genetic markers that characterize individuals (Gineikiene et al., 2009). Those PCR-based molecular markers used for chimerism evaluation includes SNP, Indel and tandem repeats. Because of their relative stability and uniqueness, they are proved to be particularly useful as molecular markers for monitoring chimerism and can be analyzed by sensitive quantitative methods (Gineikiene et al., 2009; Oliver et al., 2000). Also due to the convenience of distinguishing majority genotypes, in previous studies on human, the Indel polymorphism-based qPCR system was successfully used for genotyping recipient/donor pairs for chimerism assessment (Alizadeh et al., 2002; Gineikiene et al., 2009).

By analyzing pedigree and separating genotypes of each cell layer by somatic embryogenesis or progeny analysis, the chimera configuration can be established (Moncada et al., 2006). In order to analyze the genetic composition of different cell layers of a chimeric plant, reliable polymorphic markers distinguishing mutants from parent cultivars are required. DNA markers such as SSR and RAPD, biochemical components, morphological markers, cytological markers, isozymic markers and gene expression patterns have been successfully applied for identification of different genotypes of donor

plants and intercellular interactions in previous chimeric plants studies (Zhang et al., 2007). Induced ploidy differences between cells in different apical layers have been used to analyze development of organs in relation to apical cell lineage of a periclinal chimera in datura (SATINA, 1944; SATINA, 1945; SATINA AND BLAKESLEE, 1941; SATINA et al., 1940; SATINA AND BLAKESLEE, 1943). In citrus, the structure of graft chimeras was identified by analyzing peroxidase isozyme, RAPD and SSR markers that were specific to tissues of leaf and fruit from both donor plants (APRILE et al., 2011; SUGAWARA et al., 1995; SUGAWARA et al., 2002; ZHOU et al., 2002). The naturally occurring 'Zaohong' navel orange is a periclinal chimera consisting of cell layer I derived from Satsuma mandarin and cell layer II and III derived from 'Robertson' navel orange, with stable and valuable traits such as mixed flavor and texture from both donor plants (ZHANG et al., 2007). In a study on the synthetic periclinal chimera 'NF-5' which consists of Kawano natsudaidai and Fukuhara orange, the chemical constitution of inner side tissues of bark resembled that of Kawano natsudaidai and was distinguishable from Fukuhara orange, indicating the citrus chimera consisted of the cell layer II and cell layer III from Kawano natsudaidai and cell layer I from Fukuhara orange (OHTSU, 1994). Study of another synthetic periclinal chimera, 'NF-1', consisting of Kawano natsudaidai and Fukuhara orange, showed that the chemical constitution of juice sacs resembled that of Kawano natsudaidai and the chemical constitution of seed, segment wall, mesocarp and large vascular bundles resembled that of Fukuhara orange, indicating the chimera consisted of cell layer I from Kawano natsudaidai and cell layer II and cell layer III from Fukuhara orange (OHTSU AND KUHARA, 1994). A mirror image composition was found in 'NF-3' where the juice sac chemical constitution

resembled that of Fukuhara orange and other tissues were similar to that of Kawano natsudaidai, indicating the cell layer I was from Fukuhara orange and cell layer II and cell layer III were from Kawano natsudaidai (OHTSU AND KUHARA, 1994). The studies of Kobayashi Mikan, including peroxidase isozyme pattern analysis of leaves and nucellar seedlings, scanning electron microscopy of the epidermal system of leaves, pollen and fruits, as well as isozyme pattern analysis of juice sac, flavedo, albedo and seedlings, concluded that Kobayashi Mikan is a true periclinal chimera consisting of cell layer I from satsuma mandarin and cell layer II from Natsudaidai (OHTSU AND KUHARA, 1994). Citrus fruit sector chimeras have been observed in several citrus cultivars. Iwamasa et al. found in a fruit sector chimera of 'Fukuhara' orange, that the seeds produced within the yellow rind sector gave rise to plants with fruit with entirely yellow rinds, while the seeds produced within the normal orange rind sector gave rise to plants with fruit bearing entirely normal orange rinds (BOWMAN et al., 1991; IWAMASA et al., 1977). Also, the sectored chimeric fruits have been observed to have a distinctly different susceptibility to rind damage by diseases or pests between different sectors (BOWMAN et al., 1991). Some fruit sector chimeras have different genotypes between flesh and rind, and this may due to the mutation being restricted to only one cell layer (ZHANG et al., 2007). Therefore the chimera demonstrates genetic mutations or somatic segregation or spontaneous polyploidization, and contributes a potentially valuable source of genetic variants for citrus breeding (BOWMAN et al., 1991).

Plant chimeras that carry distinct phenotype markers are valuable for development studies because the tissues and organs produced in different cell layers can be investigated by

tracing the fate of apical cells. The generation of plant chimeras, the cell division pattern and cell fate during plant development can be deduced from analysis of plant chimeras, because the patterns of division and differentiation of derivative cells rely on the position of the original cells in the meristem (SZYMKOWIAK AND SUSSEX, 1996). Phenotypic markers were used to study the genetic origin of cells and tissue development in periclinal chimeric tomato (SZYMKOWIAK AND IRISH, 1999; SZYMKOWIAK AND SUSSEX, 1992). By using periclinal chimera tobacco with variegated leaf, the origin of eggs was found to be from cell layer II (MARCOTRIGIANO AND BERNATZKY, 1995). By monitoring genetic markers, the parental sources of the meristem layers can be identified. For instance, in the citrus studies of two synthesized periclinal chimeras from 'Hamlin' orange and 'Satsuma' mandarin and two naturally occurring chimeral cultivars, Kobayashi Mikan and Kinkoji Unshu, RAPD markers showed that the outermost layer (layer I) of the shoot apical meristem consisted of cells from a different species from that in the inner layers (layer II and layer III) of the four graft chimeric cultivars, and the juice vesicles are derived from not only from cell layer I but also from sub-epidermal cells, which derive from cell layer II (SUGAWARA et al., 2002). The segment membrane is derived from adaxial epidermal cells of endocarp and cells from inner tissue under the epidermis, which means the segment membrane develops from cell layer I and cell layer II, and possibly cell layer III (SUGAWARA et al., 2002). In contrast, the fruit albedo tissue is derived from a single genotype and does not contain the epidermal tissue of the pericarp, and multiple nucellar embryos were derived from the same layer as mesophyll (SUGAWARA et al., 2002). In addition to morphological and genetic markers, marker genes have also been used to track

cells in specific cell layers of meristem, and their descendants, over developmental time. Marker genes disrupted by ionizing radiation or active transposons produce mutated cells that are genetically different from those in the rest of the plant body, and mutants that can be visually detected, such as genes in pigmentation pathways, are particularly useful (Frank and Chitwood, 2016). Analysis of GUS expression and anthocyanin pigmentation also verified that leaf structure consists of cells derived from all three cell layers of meristem (Kan et al., 1996). By utilizing SNPs and RNA sequencing, L1-derived transcripts and L2/L3-derived transcripts were distinguished by SNPs that differentiate *S. pennellii* and *S. lycopersicum* in the chimera between *S. pennellii* (L1) and *S. lycopersicum* (L2/L3) (Filippis et al., 2013). The RAPD markers used in fruit structure analysis of interspecific periclinal chimeric citrus composed of different cultivars in L1 and L2/L3 showed that the albedo and embryos produced the same product as in mesophyll in each chimera, while other fruit tissue, such as segment membrane, juice vesicle and pericarp, produced products corresponding to both parent cultivars (Sugawara et al., 2002).

However, due to the high plasticity and totipotency of plant cells, it is possible that the cell from one cell layer can invade the adjacent cell layer and subsequently adopt the cellular identity and eventually change its development trajectory to that of the adjacent cells during the later stage of differentiation (FRANK AND CHITWOOD, 2016). Based on this, Irish and Sussex proposed the term "probability mapping", in which position is more likely than the lineage to determine the fate of plant cells even late in development (FRANK AND CHITWOOD, 2016; IRISH AND SUSSEX, 1992).

In general, chimerism not only serves as an advantageous tool for plant developmental studies but also is beneficial for understanding intercellular communication, context-dependent gene function, cell lineages during development, and the consequences of heterogenomicity within an individual (FRANK AND CHITWOOD, 2016).

In this study, the chimerism of the Tango mandarin is investigated by comparing the genetic composition between Tango tissues and to its parent, W. Murcott mandarin. W. Murcott mandarin, also known as Afourer, and perhaps as Nadorcott is originally from Morocco (WU et al., 2014). W. Murcott is a complex hybrid found as a chance zygotic seedling from the Murcott tangor and an uncertain pollen parent, and Murcott itself is presumed an F1 hybrid of a cross between an unknown mandarin and a sweet orange (WU et al., 2014). The potential grandparent-grandchild relationship between sweet orange and W. Murcott was supported by sequence analysis, which showed 34% of the sweet orange genome is shared with W. Murcott (WU et al., 2014). W. Murcott was imported to California in 1985 and has been widely planted throughout California since then due to its high-quality fruits, which are deep orange in color, sweet juicy and easily peelable (ROOSE AND WILLIAMS, 2007). One of its valuable characters is that W. Murcott fruits can be seedless because W. Murcott is self-incompatible, which means that its female gametes are unable to be fertilized by its male gametes, preventing seed formation and resulting in facultative parthenocarpy (GAMBETTA et al., 2013; OLLITRAULT et al., 2007). The setting of seedless fruits of W. Murcott mandarin requires that trees be either planted in isolated areas or are covered with the screen during the flowering season, otherwise, they will produce seedy fruits of lesser value as a result of cross-pollination by bees-transferring

pollen from other Citrus with high pollen viability (CROWLEY, 2011). Tango mandarin is a low-seeded selection derived from gamma irradiation of budwood of diploid cultivar W. Murcott. Budwood of W. Murcott was treated with 50 Gray units of ^{60}Co gamma at Riverside, California and then propagated on various rootstocks in a greenhouse (ROOSE AND WILLIAMS, 2007). During initial evaluation of this population, one tree propagated on 'C32' citrange rootstock showed distinctive characteristics, specifically that fruits were set with very low seed content in all situations of cross-pollination and fruit quality and production were the same as the original 'W. Murcott' cultivar (ROOSE AND WILLIAMS, 2007). This tree was later named 'Tango' and selected for further trials. Buds derived from this tree were propagated onto 'Carrizo' and 'C35' citrange rootstock and evaluated for disease status for multiple rounds before 'Tango' was commercially released (ROOSE AND WILLIAMS, 2007).

The radiation may cause DNA double-strand breaks (DSBs), the following repair of which is error-prone and may cause chromosome structure changes, including translocations, inversions and deletions, and local nucleotides changes such as SNPs and Indels (SANKARANARAYANAN et al., 2013). According to a previous observation of homologous chromosome misalignment, chromosome segregation failure and lagging chromosomes during meiosis of Tango (CROWLEY, 2011), it is considered that gamma irradiation caused structural variations in cells that give rise to gametes. Also based on the observation of low germination rate and small size of pollen, and setting low-seeded fruits in a cross-pollination condition, both male and female gametophytes are likely affected by mutations resulting from gamma irradiation. Therefore, it is plausible that gamma irradiation initially

affected one or a few cells of cell layer II positioned near the apical dome of the shoot apical meristem, eventually making Tango a periclinal chimera with at least cell layer II carrying the mutation. The genomic composition of layers I and III may have remained the same as in W. Murcott. The objective of this study is to investigate whether the previously discovered structural variations are present in cell layer II and/or other cell layers of Tango, in order to better interpret the previous observations of abnormal chromosome pairing and segregation during meiosis and to better predict genetic stability of Tango. PCR primers distinguishing the structural variations in Tango from the sequences present in W. Murcott were used to amplify DNA from different tissues of Tango.

Materials and Methods

Tissues from different organs of Tango were collected and genotyped using primers previously identified as distinguishing Tango leaf DNA from that of W. Murcott. Since Tango mandarin is derived from gamma irradiation of W. Murcott mandarin and could be a chimera, these genomic differences may be reflected in one or more tissues while other tissues of Tango are the same as W. Murcott. Primers previously verified in Chapter II and Chapter III as detecting genomic differences between Tango and W. Murcott were applied to DNA from different tissues of Tango.

Fruit Collection

Tango mandarin fruits were collected from five randomly selected trees from a Tango planting in field 10K at UCR. Fruits were stored in a cold room before analysis.

Tissue Isolation

Fruits were placed on ice during tissue separation. For each fruit, flavedo was firstly removed, then albedo were sliced carefully with a clean blade, avoiding inclusion of other tissue as much as possible, then packed into aluminium foils and placed on dry ice respectively. Juice vesicles were separated from segment membranes with a clean blade,

and juice vesicles from the same fruit were pooled into a 50 ml centrifuge tube and placed on dry ice. All samples were stored in the -80 ˚C freezer after tissue isolation.

Juice Vesicle Tissue Dehydration

Because juice vesicles contain a large proportion of water, in order to efficiently extract DNA, the juice vesicles were freeze-dried before DNA extraction. A lyophilizer was started two days before using to let it completely cool down. The cap of each 50 ml centrifuge tube containing juice vesicles was removed, and the tube opening was covered with a double-layered Kimwipe fixed with a rubber band. 50 ml centrifuge tubes containing juice vesicles were placed on dry ice before transfer to vacuum flasks.

The vacuum flask was placed in an ice bucket and packed with dry ice between vacuum flask and ice bucket to keep it cool. 50 ml centrifuge tubes containing juice vesicle samples were carefully put into the vacuum flask with their opening facing up. Vacuum flask and glass connector were connected to lyophilizer chamber and sealed with vacuum grease. The samples were dried completely on the lyophilizer before being stored in the -80 ˚C freezer.

Tissue Grinding

Before grinding of each tissue, mortar and pestle were cooled with liquid nitrogen. Tissues collected were placed into the mortar with liquid nitrogen, then ground with a pestle into

fine powders. Powders were collected into 1.5 ml centrifuge tubes or 15 ml centrifuge tubes depending on the volume of powder.

DNA Extraction

For DNA extraction, ~100mg frozen tissue was extracted using Qiagen DNeasy Plant Mini Kit following the manufacturer's instructions except at step 19 where instead of pipetting another 100 µl Buffer AE onto the DNeasy membrane, ~100 µl of the solution that had just passed through the membrane was added, and centrifuged into a microcentrifuge collecting tube, in order to enhance the concentration of extracted DNA. After DNA extraction, the DNA concentration was measured with a Nanodrop 2000 spectrophotometer. This DNA concentration in ng/µl was used to calculate the amount of DNA used in the qPCR reaction.

Pollen Collection

Tango flower buds were collected from different Tango trees in field 10K at UCR during the flowering season. Buds were collected when size and morphology indicated that they would open within 24 hours. For each flower bud, the petals were removed and anthers were carefully picked without filaments using fine point forceps and dispersedly in a clean Petri dish. Then the Petri dishes were placed under an incandescent desk lamp within about 20 cm from the bulb for overnight in order for the anthers to dehisce. Each anther was held with clean fine point forceps and placed in a 1.5 ml centrifuge tube containing 30 ul

Millipore water, followed by being gently shaken several times in the 1.5 ml centrifuge tube before the anther was taken out. This process was repeated with flowers from the same tree and the pollen samples were combined in the same 1.5 ml centrifuge tube to increase the yield. The 1.5 ml centrifuge tube was then centrifuged at 5000 rpm for 10 min and the supernatant was discarded, and the pellet was dried at room temperature for 5 min before pollen was collected. The purity of collected pollen was determined by examining a sample under a Carl Zeiss binocular optical microscope (Model 47 30 11-9901) with 400X magnification to avoid inclusion of other plant tissue. Pollen DNA extraction procedure was performed immediately after pollen was collected.

Pollen DNA Isolation

The pollen DNA extraction followed the methods by Lalhmangaihi et al. (2014). Pollen was suspended in 500 µl extraction buffer (100 mM Tris-HCl, 50 mM EDTA, 50 mM NaCl, 10% SDS, pH 7.5) in the 1.5 ml centrifuge tube used for pollen collection followed by transfer to a 2 ml flat-bottom tube. Sterilized glass beads 0.5 g (diameter 0.5–1 mm) were added to the 2 ml tube and pollen was ground with a plastic pestle for 5–10 min. 100 µl DTT (110 mM) and 10 µl proteinase K (10 mg/ml) were added to the tube and mixed by gentle inversion followed by incubation at 56°C water bath for 1 hour. Then 500 µl CTAB extraction buffer (20 mM Tris-HCl, pH 8.0, 10 mM EDTA, pH 8.0, 10% CTAB, 5% polyvinylpyrrolidone), 10 µl proteinase K (10 mg/ml), and 50 µl DTT (110 mM) were added to the tube and the mixture was incubated at 65°C overnight in water bath. 500 µl

phenol-chloroform-isoamyl alcohol (25:24:1 v:v:v) was then added and centrifuged at 10,000 rpm for 10 min. The supernatant was transferred to a new 2 ml centrifuge tube, 500 µl isopropanol and 100 µl sodium acetate (3 mM) were added and kept at −20°C for 1 h for precipitation followed by centrifugation at 13,000 rpm at 4℃ for 10 min. The supernatant was transferred into a new 2 ml centrifuge tube, followed by mixing with 400 µl 70% ethanol, and incubation at −20°C for 15 min. Then the mixture was centrifuged at 13,000 rpm for 10 min at 4℃, and the supernatant was pipetted off carefully without disturbing the pellet. The pellet was dried in an oven at 50 °C followed by adding 30 µl Millipore water to the tube and mixing gently.

Real-Time PCR Primer Design for Tango Tissue Indel Markers

In order to verify Indels short as 1-3 bp with high resolution, real-time PCR was utilized instead of conventional PCR and the primers were designed with the same methods as in Chapter II. One primer set was finally selected for each Indel locus, seven in total, for Tango tissue-specific analysis (Table 4.1).

Real-Time PCR Amplification for Tango Tissue Indel Analysis

Real-time PCR was performed and analyzed with ABI QuantStudio. The reaction mixture contained diluted DNA, 1X MeltDoctor HRM Master mix and 10 uM of each primer in a total volume of 20 ul. The real-time PCR cycle profile was: 10 min of an activation period

at 95 ℃, followed by 40 cycles of a two-step profile involving 15 sec at 95 ℃ for denaturation and 1 min at 60 ℃ for annealing and extension. The melt analysis conditions were 10 sec at 95 ℃, followed by 1 min at 60 ℃, and a ramp from 60 ℃ to 95 ℃, rising by 0.25 ℃ every 5 sec, and a final 15 sec at 60 ℃. Five albedo DNA samples and five juice vesicle DNA samples were used in this analysis, and two Tango leaf DNA samples and two W. Murcott leaf DNA samples were used as reference. Melt curves and derivative melt curves were plotted for each pair of primers using Applied Biosystems QuantStudio™ 6 & 7 Flex Real-Time PCR System.

Real-Time PCR Primer Design and PCR for Chromosome Structural Variant Detection

In order to verify the presence/absence of chromosome rearrangement breakpoints in different tissues of Tango, primer sets whose products include the breakpoints were designed with the same methods as in Chapter III, in which the sequences of reads affected by the same structural variant were first assembled into longer contigs and primer pairs were designed from contig sequences to span each breakpoint. Multiple primer pairs were designed in a similar way (Table 4.2), and an MDH gene primer pair Cit1249 and Cit1250 (sequence unpublished) was used as positive control. Real-time PCR preparation procedure and condition were the same as in Chapter III. Seven Tango pollen DNA samples in total were analyzed.

Relative Quantification Analysis

Since the Tango tissue samples analyzed could have been contaminated with tissue from adjacent cell layers during tissue isolation and this could be detected by sensitive real-time PCR, the relative quantification method was used to determine the approximate copy number of the novel sequences (i.e., the chimeric sequences which result from structural variations) in different Tango tissues. Instead of absolute quantification of the physical amount of target sequence, comparing the Ct value of the target sequence to an endogenous control gene is the best practical way to determine the relative copy number of two or more targets. To quantitatively compare the copy number of two or more targets, one of the most robust methods by real-time PCR is the comparative Ct method, which is also called delta-delta Ct method. In this method, the difference of Ct value (ΔCt) between the Ct value for target sequence (Ct_t) and the Ct value for the endogenous control (Ct_e) of the same DNA sample is independent of DNA concentration (BUBNER AND BALDWIN, 2004):

$$\Delta Ct = Ct_t - Ct_e$$

In addition, the ratio of the amount of two target sequences can be simply calculated as follows:

$$\frac{X_{t_1}}{X_{t_2}} = 2^{-\Delta\Delta C_t}$$

where

$$\Delta\Delta Ct = \Delta Ct_1 - \Delta Ct_2$$

(Bubner and Baldwin, 2004). In this part, the MDH gene primer pair was used as the control for quantifying target sequence copy number because it is single copy within the haploid reference genome and has been sequenced in many citrus varieties (RAMADUGU et al., 2013). Since fruit albedo tissue is completely derived from cell layer II of meristem, and juice vesicles are completely derived from cell layer I of the meristem, those two tissues can be used to represent the genomic composition of two cell layers due to their relative ease of isolation from other tissues. The real-time PCR was performed three times for each sample and Ct values were recorded for calculation. For each sample, mean Ct value of the product of each structural variation primer set and mean Ct value of the product of MDH primer set were calculated, the differences between which were averaged as ΔCt.

For each sample: $$\Delta Ct_{SV} = \overline{Ct_{SV}} - \overline{Ct_{MDH}}$$

For each structural variant, the statistical significance of ΔCt between albedo and juice vesicle was performed using independent two-sample Student's t-test with unequal sample sizes and equal variance. Then ΔCt_{SV} across all samples of the same tissue were averaged for $\Delta\Delta Ct$ calculation.

Results

Since the Indel markers previously identified were relatively small and conventional PCR and separation of products on agarose gels did not provide sufficient resolution, real-time PCR was utilized to determine which tissue samples contained the Tango-specific mutations. Because gamma irradiation could affect only one or a few cells in a shoot apical meristem, it is most likely that cell layer II was affected since this layer gives rise to gametes, and Tango showed abnormal chromosome pairing and separation during meiosis observed in previous studies (CROWLEY, 2011). The other two cell layers may not be affected by gamma irradiation or may not carry the same mutations as Tango, so for these markers the genomic composition of other layers may still be the same as in W. Murcott. It is also possible that cells in other cell layers have been invaded and replaced by the affected layer II cells, resulting in a genomic composition that is uniform through all cell layers for the targeted mutation. Therefore, the genomic differences of Tango from W. Murcott could reside only in cell layer II or another cell layer as well. Tissues chosen for analysis were expected to derive mostly or entirely from single cell layers, and be conveniently separable from tissues composed of cells from other layers.

Different Melt-Curve Patterns of Indel Loci between Tango Fruit Tissues

Tango albedo and juice vesicle tissue were analyzed for seven Indel markers (DC-1, DC-7, Del2, Del3, Del4-2, Del3-3 and Del2-3) which differentiated Tango leaf DNA from W.

Murcott leaf DNA. For all seven primer sets, both Tango leaf DNA and W. Murcott leaf DNA showed uniform melt-curve patterns which differed between varieties, and no products were detected for NTC (no template control). All five Tango albedo samples produced the same products as two Tango leaf DNA samples and differed from products obtained with two W. Murcott leaf DNA samples. For juice vesicle DNA samples, five out of seven Indel markers (DC-7, Del2, Del4-2, Del3-3 and Del2-3) showed the same melt-curves as the products of Tango leaf DNA. However results with primers for DC-1 and Del3, showed products of juice vesicle DNA were similar to those from W. Murcott leaf DNA but contained some signal from products with a melt-peak similar to that observed with Tango leaf DNA. For DC-1, W. Murcott leaf DNA samples had a major peak at approximately 72.5 ℃, and Tango leaf and Tango albedo DNA samples had a major peak at approximately 71 ℃, while Tango juice vesicle DNA samples had a major peak at approximately 72.5 ℃ with a shoulder at approximately 71 ℃. For Del3, W. Murcott leaf, Tango leaf, Tango albedo and Tango juice vesicle DNA samples all had a major peak at 75 ℃, while the minor peaks of W. Murcott leaf and Tango juice vesicle DNA samples were at approximately 74 ℃, compared to the minor peaks of Tango leaf and Tango albedo DNA samples at approximately 73 ℃. Also, we observed that the minor peaks of Tango juice vesicle DNA samples were broader than the minor peaks of W. Murcott leaf DNA samples (Figure 4.1 - 4.7).

HRM Analysis Shows Three Chromosome Rearrangements Present in Most Tango Pollen Samples

Pollen grains were examed under binocular microscope and they were found to be scattered without other tissue observed (Figure 4.8). All the Tango pollen DNA samples were amplified with the MDH primers, and NTC did not amplify with these primers, indicating that DNA was successfully extracted from pollen (Figure 4.9 - 4.15 pink curves). The results of real-time PCR of Tango pollen with different primer sets are summarized in Table 4.3.

Translocation TRA1/3:

For primer pair NB13 and NB14 which target a putative translocation involving chromosomes 1 and 3, the melt peak of real-time PCR showed that four out of seven Tango pollen DNA samples were amplified by the translocation primers (red curves) with the same product as the positive control (Tango leaf DNA), while the other three Tango pollen DNA sample failed to amplify (Figure 4.9).

Deletion DEL2:

For the large deletion DEL2 on chromosome 2, there are two local assemblies present in Tango and absent in W. Murcott, which correspond to different breakage-rejoining events. The melt peaks of Tango pollen samples had nearly the same melt peaks as in the Tango leaf DNA (positive control) with melting temperature range within 1 ˚C, indicating the amplification of Tango pollen DNA produced the same products as in Tango leaf DNA, which are the sequences joined at the breakpoints that are only present in Tango and absent

in W. Murcott. Three primer sets were thus designed targeting a putative deletion or translocation on chromosome 2 according to the two local assemblies, in which NB81 + NB82 primer pair and NB87 + NB88 primer pair are expected to amplify DEL2 local assembly 1 and NB93 + NB94 primer pair are expected to amplify DEL2 local assembly 2. For primer pair NB81 and NB82, the melt peak of real-time PCR showed that all Tango pollen DNA samples were amplified by the deletion primers (red curves) with the same product as the positive control (Figure 4.10). For primer pair NB87 and NB88, the melt peak of real-time PCR showed that six out of seven Tango pollen DNA sample were amplified by the deletion primers (red curves) with the same product as positive control, while one Tango pollen DNA sample failed to amplify (Figure 4.11). For primer pair NB93 and NB94, the melt peak of real-time PCR showed that all Tango pollen DNA samples were amplified by the deletion primers (red curves) with the same product as the positive control (Figure 4.12).

Inversion INV4:

For the large inversion INV4 on chromosome 4, most Tango pollen DNA samples produced nearly the same product as in Tango leaf DNA (positive control) with melting temperature ranging within 1 ˚C, which were the breakpoint-joined sequences that are present in Tango and absent in W. Murcott, except a few samples failed to amplify. There are three primer sets designed targeting a putative inversion on chromosome 4 and utilized according to the two local assemblies of INV4, where NB143 + NB144 primer pair for INV4 local assembly 1 and NB115 + NB116 primer pair and NB153 + NB154 primer pair for INV4 local assembly 2. For primer pair NB143 and NB144, the melt peak of real-time

PCR showed that six out of seven Tango pollen DNA samples were amplified by the inversion primers (red curves) with the same product as the positive control (Tango leaf DNA), while one Tango pollen DNA sample failed to amplify (Figure 4.13). For primer pairs NB115 and NB116, the melt peak of real-time PCR showed that six out of seven Tango pollen DNA samples were amplified by the inversion primers (red curves) with the same product as the positive control (Tango leaf DNA), while one Tango pollen DNA sample failed to amplify (Figure 4.14). For primer pairs NB153 and NB154, the melt peak of real-time PCR showed that five out of seven Tango pollen DNA samples were amplified by the inversion primers (red curves) with the same product as the positive control (Tango leaf DNA), while the other two Tango pollen DNA sample failed to amplify (Figure 4.15).

Three Chromosome Rearrangements Are Single Copy Sequences in Albedo

The mean ΔCt values of each chromosome rearrangement for both Tango albedo and juice vesicle tissues were calculated, and differences between tissues in mean ΔCt were tested using the independent two-sample Student's t-test with unequal sample sizes and equal variance for differences (Table 4.4). One primer set, NB13 and NB14, was used to amplify the junction sequence of TRA1/3. Three primer sets, NB81 and NB82, NB87 and NB88, and NB93 and NB94, were used to amplify the junction sequences of DEL2. Three primer sets, NB143 and NB144, NB115 and NB116, and NB153 and NB154, were used to amplify the junction sequences of INV4. Most mean ΔCt values of albedo with structural variation primers were around 1, except primer sets NB87 + NB88 and NB143 + NB144, indicating

the copy number of sequences involved in these structural variations with respect to MDH gene in the albedo tissue is around:

$$\frac{CN_{SV}}{CN_{MDH}} = 2^{-\Delta\Delta Ct} = 2^{-1} = 0.5$$

since $\Delta Ct_{MDH} = 0$ with respect to itself, which means the copy number of structural variant sequences is half of the copy number of MDH sequences in a diploid genome. The mean ΔCt values of juice vesicles with structural variation primers were around 2.5, except primer set NB143 + NB144, and therefore the ratio of copy number of sequences involved in structural variation relative to the MDH gene is around:

$$\frac{CN_{SV}}{CN_{MDH}} = 2^{-\Delta\Delta Ct} = 2^{-2.5} \approx 0.18$$

where $\Delta Ct_{MDH} = 0$ with respect to itself.

The independent two-sample Student's t-test showed that the p-values were less than 0.05 for all seven structural variation primer sets, indicating that copy number of target sequences (structural variation sequences) were significantly different between albedo and juice vesicles, with albedo having 2.6 to 3.4 fold higher copy number than juice vesicles (Figure 4.16).

Discussion

Chimerism is widely discovered in citrus varieties and other species. 'Zaohong' navel orange, a naturally occurring grafted periclinal chimera from a 'Robertson' navel orange bud top-worked onto Satsuma mandarin, has a genomic composition where cell layer I is similar to Satsuma mandarin while cell layers II and III are similar to Robertson navel orange (ZHANG et al., 2007). In mutation breeding of potato, a periclinal chimera was found with layer I differing from cell layers II and III, which was reflected in the tuber skin color where cell layer I mutated from red towards anthocyanin-free, resulting in small red patches (VAN HARTEN et al., 1981). In a study of the fleshless mutation in grapevine, the mutation was considered as a single dominant allele located only in cell layer II, the interaction of which with wild-type cell layer I within the chimera resulted in a fleshless berry (FERNANDEZ et al., 2006). Both cell layers I and II of mutant-phenotyped offspring from selfing carried the mutated allele and this lead to fruit set failure (FERNANDEZ et al., 2006).

Cell displacement can also occur, where cells in the outer layer take over the position of cells from the inner layer and make the tissues derived from the inner layer express characters of the outer layer, or vice versa, and the inner tissues and epidermis tissues could impose effects on each other, during leaf and fruit development (MARCOTRIGIANO, 1990; ZHOU et al., 2002). Cell displacement in a chimeric meristem could eliminate chimerism and further change the genotype of cell layers. Periclinal cell divisions in the epidermal

layer followed by further anticlinal cell division can result in the appearance of sectors or patches.

Qualitative Analysis of Indels in Tango Fruit Tissue

Seven previously verified Indel markers were utilized for Tango fruit tissue analysis. The result showed that for DC-7, Del2, Del4-2, Del3-3 and Del2-3, Tango albedo and Tango juice vesicles showed the same products as in Tango leaf samples, and different from products of W. Murcott leaf samples, indicating these five Tango-specific markers could not differentiate Tango juice vesicle tissue from Tango albedo tissue. In contrast, for markers DC-1 and Del3, the products of Tango juice vesicles were different from those of Tango albedo, and similar to those of W. Murcott leaf tissue but contain some products typical of Tango leaf tissue, indicating the two markers could differentiate Tango juice vesicle tissue from Tango albedo tissue. Since the juice vesicles are derived, at least mainly, from cell layer I, and albedo is derived from cell layer II, the amplification from these two tissues may indicate that Tango is chimeric with respect to DC-1 and Del3 markers but genetically uniform with respect to DC-7, Del2, Del4-2, Del3-3 and Del2-3 markers within these two cell layers. Illumina sequencing and Sanger sequencing data of leaf tissue, which consists of cells from all three cell layers of shoot apical meristem, showed all seven short deletions are present in at least one cell layer of the shoot apical meristem in Tango while absent in W. Murcott. Based on the above results, it is likely that the short deletions of DC-7, Del2, Del4-2, Del3-3 and Del2-3 are present in both cell layer I and cell layer II, while

short deletions DC-1 and Del3 are only present in cell layer II but absent in cell layer I of Tango shoot apical meristem.

Based on the above analysis, in order to investigate whether juice vesicle tissue is derived only from cell layer I or both cell layers, a practical method is TA cloning of PCR products of juice vesicle tissue with the Indel primers followed by analyzing whether the proportion of clones whose sequence carries the short deletion is close to 50%. However, this method would be rather laborious to apply because it requires sequencing a relatively large number of clones from different trees or fruits to determine whether all or a portion of cells carry short deletions. Another improvement could be relative quantification of PCR products of the Indel primer sets on different tissues in parallel with the product of a housekeeping gene with known copy number in the genome as internal control and compare the difference of their ΔCt between different tissues, where the ΔCt is defined in the same way as the difference of real-time PCR amplification Ct value of a sample with the Indel primer set and the housekeeping gene primers.

Chromosome Rearrangements in Tango Pollen DNA

The amplification of Tango pollen DNA with MDH primers indicated that genomic DNA was successfully extracted from the pollen, but for some samples, the real-time PCR reactions with structural variation primers failed. For the suspicious translocation TRA1/3, nearly half of Tango pollen DNA samples failed to amplify. The reason could be that this rearrangement is not present in all pollen samples or the samples that failed to amplify with

TRA1/3 primers only have a very small portion of the grains carrying the rearrangement, which became undetectable. Thus this structural variant could be an unstable mutation, and the Tango trees sampled were mericlinal or sectorial chimera with respect to the mutation. However, the actual structural variant has not yet unambiguously identified, and may be caused by activity of a transposable element or be a duplication rather than a translocation event. Therefore it is challenging to conjecture a plausible cause for the relatively higher rate of amplification failure. In contrast, for the large deletion DEL2 on chromosome 2 and the large inversion INV4 on chromosome 4, most the pollen samples produced PCR products with nearly identical melt peaks (range about 1 ˚C).

The failure of amplification of some Tango pollen samples with structural variant primer sets could result from one or more following reasons: 1. Tango is heterozygous for these rearrangements, in which the targeted rearranged sequences have only one copy while the MDH gene has two copies in the diploid genome, so that MDH template concentration is adequate for amplification but rearrangement sequence is not. 2. The meiotic products that carry one or more chromosome rearrangements may not be viable and are finally aborted during pollen development, leading to the majority of collected pollen grains being normal haploids without chromosome rearrangements. This is also supported by the observation of almost uniform size and shape of pollen grains under microscope (Figure 4.8) and previous observation that most Tango pollen grains are smaller than those of W. Murcott and have low viability (ROOSE AND WILLIAMS, 2007). The method used for pollen isolation discriminated against pollen which did not complete a normal meiosis so that DNA was isolated mostly from pollen lacking the rearrangement. 3. The expected proportion of

abnormal pollen depends on the size of the rearrangement and the probability of crossovers in the affected region. Dicentric chromosome or acentric chromosomal fragment may be produced by crossovers within the rearrangement and fragments which contain the target chimeric sequences are eventually lost at telophase of meiosis. The above possibilities reduced the chance of rearranged DNA sequence giving PCR products with the rearrangement-specific primers.

Quantitative Analysis of Chromosome Rearrangements in Tango Fruit Tissue

The relative quantitative analysis of copy number of sequences that were affected by structural variation showed that the ΔCt between albedo and juice vesicles had significant differences ($p < 0.05$) for all primer sets of three putative structural variations, indicating differences in the copy number of chimeric sequence within cells from albedo and juice vesicles (Table 4.4). The copy number of structural variations of albedo is about half of that of MDH, as expected for comparison between a diploid with two copies of MDH and a sample which is hemizygous for a novel single copy sequence. The copy number of structural variants in juice vesicles is near one fifth of that of MDH, which is not consistent for with expected ratios for homogeneous populations of cells homozygous or hemizygous for the variant. Also, for chimeric sequences involved in structural variations the copy number ratios of albedo/juice vesicles were larger than 2.5 (Table 4.4), indicating that the copy number of these sequences in juice vesicles is less than 40 % of that in albedo. Therefore, the amplification of these sequences from juice vesicle DNA may result from

these samples containing a small proportion of DNA from cells derived from cell layer II. Suguwara also suggested that juice vesicle tissue contains some cells from layer II, which could also be responsible for the small amount of amplification of juice vesicle DNA (SUGAWARA et al., 2002).

According to the results above, it is verified that the DNA extracted from many Tango pollen grains carries the structural variant mutations. Since gametes are developed from cell layer II of the shoot apical meristem, it is likely that the cell layer II was affected by the gamma radiation and the tissues derived from cell layer II also carry the chromosome rearrangements. Since pollen is haploid, the aneuploidy resulting from meiotic crossovers in a structural rearrangement heterozygote could bring about inviable and infertile pollen, and further cause low-seeded fruit. The genomes in cell layer I may not carry these rearrangements, since the relative copy number of structural variants in juice vesicle DNA, which is thought to be derived from cell layer I, is below the normal copy number range and may result from a mixture of cell layers in the tissue sampled.

Possible Occurrence of Mutation Events

The Illumina sequencing of W. Murcott used a leaf sample from a W. Murcott tree in the Citrus Variety Collection at University of California, Riverside, which was propagated from one of several W. Murcott source trees maintained by the UCR Citrus Clonal Protection Program (CCPP). The Illumina sequencing of Tango used a leaf sample from a Tango tree grown at UCR field 1B, which was propagated directly from the Tango mother

tree. The Tango mother tree was generated by irradiation of buds from one of several W. Murcott source trees of the CCPP (Figure 4.17). It is possible that the W. Murcott tree sampled for sequencing and that which provided the budwood for irradiation which lead to Tango already had some sequence differences, possibly including some of those studied here. Additional sequence variants may have originated by natural mutation in specific cell layers after divergence of W. Murcott and Tango. Possibly gamma irradiation of the budwood of W. Murcott produced not only structural variants but also seven short deletions in one or more layer II cells, and simultaneously killed adjacent cells in the central zone of cell layer I. The replacement of cells in layer I by invasion of cells from cell layer II resulted in all structural variants and short deletions being initially carried by both cell layers. Subsequent reversion mutations of some deletions and structural variants lead to some layer I cells no longer carrying the structural variants and two short deletions, DC-1 and Del3.

To more clearly identify the propagation generation or generations in which the short deletions and structural variants emerge, the same real-time PCR procedure with Indel primers or sequencing of PCR products can be performed on leaf and fruit tissue of the W. Murcott tree that gave rise to the Tango mother tree (if it exists) and on the Tango mother tree itself, as well as the Tango tree from Lindcove that gave rise to the Tango trees in field UCR field 10K. In addition, for patent protection, the same verification can be performed on released commercial Tango trees to select markers that distinguish them from W. Murcott trees.

Overall, most of the sequence differences between W. Murcott and Tango which we detected by sequencing DNA from trees that are many cell generations and several vegetative propagation generations apart were also detected in the Tango trees sampled from field 10K and are therefore relatively stable during normal propagation. This increases our confidence in the value of the markers developed here.

Future Improvement of Mutation Visualization in Tissues

To improve cell layer analysis, several techniques could be utilized. For example, cryosection can be used for plant tissue sectioning since it provides more accuracy than manual separation and avoids the tissue deformation and potential cell-layer mixing that occurred during the room temperature slicing procedure due to melting of frozen tissue. In order to better identify the tissue or tissues that contain the chromosome rearrangements, multiplex fluorescence in situ hybridization (M-FISH) could be utilized with probes that specifically hybridize to the regions flanking the breakpoint of each chromosome rearrangement using different fluorescence dyes. By microscopic visualization and digital acquisition of images of mitotic metaphase of cells from different tissues, the adjacency of the fluorescence signals that are supposed to flank the breakpoint on the same chromosome would indicates the presence of chromosome rearrangements in the target cell, while the separation of the fluorescence bands on different chromosomes or in different regions of the same chromosome would indicate the absence of chromosome rearrangements. Similarly, mutaFISH™ (mutation-specific fluorescence in situ hybridization) could be

utilized to specifically determine the presence or absence of the short deletions or SNPs at single nucleotide resolution (GRUNDBERG et al., 2013). By comparing the fluorescence pattern between cells from different tissues, the cell layers in which each chromosome rearrangement, SNP or short deletion is contained could be clearly identified.

References

Alizadeh, M., M. Bernard, B. Danic, C. Dauriac, B. Birebent, C. Lapart, T. Lamy, P.Y. Le Prise, A. Beauplet, D. Bories, G. Semana, and E. Quelvennec, 2002. Quantitative assessment of hematopoietic chimerism after bone marrow transplantation by real-time quantitative polymerase chain reaction. *Blood* 99:4618-4625.

Aprile, A., C. Federici, T.J. Close, L. De Bellis, L. Cattivelli, and M.L. Roose, 2011. Expression of the H+-ATPase AHA10 proton pump is associated with citric acid accumulation in lemon juice sac cells. *Functional & integrative genomics* 11:551-563.

Barton, M.K. and R.S. Poethig, 1993. Formation of the shoot apical meristem in Arabidopsis thaliana: an analysis of development in the wild type and in the shoot meristemless mutant. *Development* 119:823-831.

Boss, P.K. and M.R. Thomas, 2002. Association of dwarfism and floral induction with a grape 'green revolution'mutation. *Nature* 416:847-850.

Bowman, K.D., F.G. Gmitter, G.A. Moore, and R.L. Rouseff, 1991. Citrus-Fruit Sector Chimeras as a Genetic Resource for Cultivar Improvement. *Journal of the American Society for Horticultural Science* 116:888-893.

Bubner, B. and I.T. Baldwin, 2004. Use of real-time PCR for determining copy number and zygosity in transgenic plants. *Plant Cell Reports* 23:263-271.

Crowley, J.R., 2011. A molecular genetic approach to evaluate a novel seedless phenotype found in Tango, a new variety of mandarin developed from gamma-irradiated W. Murcott. *Ph.D. Dissertation, University of California, Riverside*.

Fernandez, L., A. Doligez, G. Lopez, M. Thomas, A. Bouquet, and L. Torregrosa, 2006. Somatic chimerism, genetic inheritance, and mapping of the fleshless berry (flb) mutation in grapevine (Vitis vinifera L.). *Genome* 49:721-728.

Filippis, I., R. Lopez‐Cobollo, J. Abbott, S. Butcher, and G.J. Bishop, 2013. Using a periclinal chimera to unravel layer‐specific gene expression in plants. *The Plant Journal* 75:1039-1049.

Frank, M.H. and D.H. Chitwood, 2016. Plant chimeras: The good, the bad, and the 'Bizzaria'. *Developmental Biology* 419:41-53.

Franks, T., R. Botta, M.R. Thomas, and J. Franks, 2002. Chimerism in grapevines: implications for cultivar identity, ancestry and genetic improvement. *Theor Appl Genet* 104:192-199.

Gambetta, G., A. Gravina, C. Fasiolo, C. Fornero, S. Galiger, C. Inzaurralde, and F. Rey, 2013. Self-incompatibility, parthenocarpy and reduction of seed presence in 'Afourer' mandarin. *Scientia Horticulturae* 164:183-188.

Gendron, J.M., J.-S. Liu, M. Fan, M.-Y. Bai, S. Wenkel, P.S. Springer, M.K. Barton, and Z.-Y. Wang, 2012. Brassinosteroids regulate organ boundary formation in the shoot apical meristem of Arabidopsis. *Proceedings of the National Academy of Sciences* 109:21152-21157.

Gillaspy, G., H. Ben-David, and W. Gruissem, 1993. Fruits: A Developmental Perspective. *Plant Cell* 5:1439-1451.

Gineikiene, E., M. Stoskus, and L. Griskevicius, 2009. Single nucleotide polymorphism-based system improves the applicability of quantitative PCR for chimerism monitoring. *The Journal of Molecular Diagnostics* 11:66-74.

Grundberg, I., S. Kiflemariam, M. Mignardi, J. Imgenberg-Kreuz, K. Edlund, P. Micke, M. Sundström, T. Sjöblom, J. Botling, and M. Nilsson, 2013. In situ mutation detection and visualization of intratumor heterogeneity for cancer research and diagnostics. *Oncotarget* 4:2407.

Hansen, M., F. Pohlheim, and C. Obermeier. 2003. Development of RAPD markers for studies on a populus-chimera, XXI International Eucarpia Symposium on Classical versus Molecular Breeding of Ornamentals-Part II 651.

Irish, V.F. and I. Sussex, 1992. A fate map of the Arabidopsis embryonic shoot apical meristem. *Development* 115:745-753.

Iwamasa, M., M. Nishiura, N. Okudal, and D. Ishiuchi, 1977. Characteristics due to chimeras and their stability in citrus cultivars. *Proceedings of the International Society of Citriculture*:571-574.

Kan, T., Y. Hirata, K. Ishiwata, S. Oguni, K. Kuno, and T. Noguchi, 1996. Cell layer constitution of interspecific chimeras in Brassica, p. 239-245. In: J.S. Dias, I. Crute, and A.A. Monteiro (eds.), International Symposium on Brassicas: Ninth Crucifer Genetics Workshop. International Society Horticultural Science, Leuven 1.

Kuhara, S., 1988. Artificial production of the synthetic periclinal chimera in citrus. *Kyushu Agric. Res* 50:222.

Marcotrigiano, M., 1990. Genetic mosaics and chimeras: Implications in biotechnology, p. 85-111, Somaclonal Variation in Crop Improvement I. Springer.

Marcotrigiano, M., 1997. Chimeras and variegation: patterns of deceit. *HortScience* 32:773-784.

Marcotrigiano, M. and R. Bernatzky, 1995. Arrangement of cell layers in the shoot apical meristems of periclinal chimeras influences cell fate. *The Plant Journal* 7:193-202.

Moncada, X., F. Pelsy, D. Merdinoglu, and P. Hinrichsen, 2006. Genetic diversity and geographical dispersal in grapevine clones revealed by microsatellite markers. *Genome* 49:1459-1472.

Ohtsu, Y., 1994. Efficient production of a synthetic periclinal chimera of citrus 'NF-5' for introduction of disease resistance. *Japanese Journal of Phytopathology* 60:82-88.

Ohtsu, Y. and S. Kuhara, 1994. Periclinal Chimera of Citrus Resistant to Citrus Canker and Citrus Tristeza Virus: Chimerism and Composition of Fruit Tissue in the Synthetic Periclinal Chimeras 'FN-1' and 'NF-3'. *Japanese Journal of Phytopathology* 60:20-26.

Oliver, D.H., R.E. Thompson, C.A. Griffin, and J.R. Eshleman, 2000. Use of single nucleotide polymorphisms (SNP) and real-time polymerase chain reaction for bone marrow engraftment analysis. *Journal of Molecular Diagnostics* 2:202-208.

Ollitrault, P., Y. Froelicher, D. Dambier, F. Luro, and M. Yamamoto, 2007. Seedlessness and ploidy manipulation. *Citrus Genetics, Breeding and Biotechnology*:197-218.

Peary, J.S., R.D. Lineberger, T.J. Malinich, and M.K. Wertz, 1988. Stability of leaf variegation in Saintpaulia ionantha during in vitro propagation and during chimeral separation of a pinwheel flowering form. *American journal of botany*:603-608.

Ramadugu, C., B.E. Pfeil, M.L. Keremane, R.F. Lee, I.J. Maureira-Butler, and M.L. Roose, 2013. A six nuclear gene phylogeny of Citrus (Rutaceae) taking into account hybridization and lineage sorting. *PLoS One* 8:e68410.

Roose, M.L. and T.E. Williams, 2007. Mandarin tree named 'Tango'. U.S. Patent No. PP17,863.

Ruth, J., E.J. Klekowski Jr, and O.L. Stein, 1985. Impermanent initials of the shoot apex and diplontic selection in a juniper chimera. *American Journal of Botany*:1127-1135.

Sankaranarayanan, K., R. Taleei, S. Rahmanian, and H. Nikjoo, 2013. Ionizing radiation and genetic risks. XVII. Formation mechanisms underlying naturally occurring DNA deletions in the human genome and their potential relevance for bridging the gap between induced DNA double-strand breaks and deletions in irradiated germ cells. *Mutation Research/Reviews in Mutation Research* 753:114-130.

Satina, S., 1944. Periclinal chimeras in Datura in relation to development and structure (A) of the style and stigma (B) of calyx and corolla. *American Journal of Botany*:493-502.

Satina, S., 1945. Periclinal chimeras in Datura in relation to the development and structure of the ovule. *American Journal of Botany*:72-81.

Satina, S. and A. Blakeslee, 1941. Periclinal chimeras in Datura stramonium in relation to development of leaf and flower. *American Journal of Botany*:862-871.

Satina, S., A. Blakeslee, and A.G. Avery, 1940. Demonstration of the three germ layers in the shoot apex of Datura by means of induced polyploidy in periclinal chimeras. *American Journal of Botany*:895-905.

Satina, S. and A.F. Blakeslee, 1943. Periclinal chimeras in Datura in relation to the development of the carpel. *American Journal of Botany* 30:453-462.

Skirvin, R.M. and M.A. Norton, 2016. 7 Chimeras. *Plant propagation concepts and laboratory exercises*:95.

Spena, A. and F. Salamini, 1995. Genetic tagging of cells and cell layers for studies of plant development. *Methods in Cell Biology* 49:331-354.

Sugawara, K., A. Oowada, T. Moriguchi, and M. Omura, 1995. Identification of Citrus Chimeras by RAPD Markers. *Hortscience* 30:1276-1278.

Sugawara, K., T. Wakizuka, A. Oowada, T. Moriguchi, and M. Omura, 2002. Histogenic identification by RAPD analysis of leaves and fruit of newly synthesized chimeric Citrus. *Journal of the American Society for Horticultural Science* 127:104-107.

Szymkowiak, E.J. and E.E. Irish, 1999. Interactions between jointless and wild-type tomato tissues during development of the pedicel abscission zone and the inflorescence meristem. *The Plant Cell* 11:159-175.

Szymkowiak, E.J. and I.M. Sussex, 1992. The internal meristem layer (L3) determines floral meristem size and carpel number in tomato periclinal chimeras. *Plant Cell* 4:1089-1100.

Szymkowiak, E.J. and I.M. Sussex, 1996. What chimeras can tell us about plant development. *Annual Review of Plant Physiology and Plant Molecular Biology* 47:351-376.

Tian, H. and M. Marcotrigiano, 1993. Origin and development of adventitious shoot meristems initiated on plant chimeras. *Developmental Biology* 155:259-269.

Tilney-Bassett, R.A., 1986. Plant chimeras. Edward Arnold (Publishers) Ltd.

Van Harten, A., H. Bouter, and C. Broertjes, 1981. In vitro adventitious bud techniques for vegetative propagation and mutation breeding of potato (Solanum tuberosum L.). II. Significance for mutation breeding. *Euphytica* 30:1-8.

Verdisson, S., F. Baillieul, and J. Audran, 2015. Use of RAPD markers to detect chimerism in synthetic grape chimeras (Vitis vinifera L.). *VITIS-Journal of Grapevine Research* 38:93.

Whitham, T.G. and C.N. Slobodchikoff, 1981. Evolution by individuals, plant-herbivore interactions, and mosaics of genetic variability: The adaptive significance of somatic mutations in plants. *Oecologia* 49:287-292.

Winkler, H., 1907. Über Pfropfbastarde und pflanliche Chimären. *Berichte der Deutschen Botanischen Gesellschaft* 25:568-576.

Winkler, H., 1909. Weitere Mitteilungen über Propfbastarde. *Zeitschrift für Botanik* I:315-345.

Wu, G.A., S. Prochnik, J. Jenkins, J. Salse, U. Hellsten, F. Murat, X. Perrier, M. Ruiz, S. Scalabrin, J. Terol, M.A. Takita, K. Labadie, J. Poulain, A. Couloux, K. Jabbari, F. Cattonaro, C. Del Fabbro, S. Pinosio, A. Zuccolo, J. Chapman, J. Grimwood, F.R. Tadeo, L.H. Estornell, J.V. Munoz-Sanz, V. Ibanez, A. Herrero-Ortega, P. Aleza, J. Perez-Perez, D. Ramon, D. Brunel, F. Luro, C. Chen, W.G. Farmerie, B. Desany, C. Kodira, M. Mohiuddin, T. Harkins, K. Fredrikson, P. Burns, A. Lomsadze, M. Borodovsky, G. Reforgiato, J. Freitas-Astua, F. Quetier, L. Navarro, M. Roose, P. Wincker, J. Schmutz, M. Morgante, M.A. Machado, M. Talon, O. Jaillon, P. Ollitrault, F. Gmitter, and D. Rokhsar, 2014. Sequencing of diverse mandarin, pummelo and orange genomes reveals complex history of admixture during citrus domestication. *Nature Biotechnology* 32:656-662.

Zhang, M., X. Deng, C. Qin, C. Chen, H. Zhang, Q. Liu, Z. Hu, L. Guo, W. Song, and Y. Tan, 2007. Characterization of a New Natural Periclinal Navel–Satsuma Chimera of Citrus:'Zaohong'Navel Orange. *Journal of the American Society for Horticultural Science* 132:374-380.

Zhou, J.M., Y. Hirata, I.S. Nou, H. Shiotani, and T. Ito, 2002. Interactions between different genotypic tissues in citrus graft chimeras. *Euphytica* 126:355-364.

Table 4.1 Summary of Real-Time PCR Primer Sets for Tango Tissue-Specific Indel Analysis

Indel	Primer Name	Sequence
DC-1	YZ_DC1_7F_SYBR	AAGAACTAGGTTGGGACCTAAA
	YZ_DC1_7R_SYBR	ACAAACATCCCTCATCTCTTAATAC
DC-7	UCR56	ACTGGGTCAGCACCCAAA
	UCR57	ACGGTGAAGGACAATTTAAGGAAAG
Del2	UCR27	ACTTAAACGTCCCGTTGCT
	UCR28	GTACCCGAAACTACGCTGTC
Del3	UCR29	TCCATTTCTGCTTCCTGTAGATAA
	UCR30	CTGGACTATAGCAGCAAGTCTAC
Del4-2	UCR35	GACGTCATGAATTACCTCTCAGAT
	UCR36	TACAAGGGCAAGATATGGCAA
Del3-3	UCR48	AATCGATCCTCAAGACCTAATGTT
	UCR49	AACGGCGTCGAAAGTGATAG
Del2-3	UCR80	CGGTGCCAAGTTGTTGTC
	UCR81	ACATCATTCAGGGCTCACA

Table 4.2 Summary of Real-Time PCR Primer Sets Amplifying Chimeric Sequences for Tango Tissue-Specific Structure Variation Analysis

Structural variation type	Assembly	Primer name	Sequence	Reference genome hits
Large Deletion DEL2	Large Deletion Local Assembly 1	NB81	CCAAGTCCTACTCCATCAATC	scaffold_2: 12563459-12563439
		NB82	TTTCCTGGGTTAAAGTGTCTAT	scaffold_2: 6572052-6572073
	Large Deletion Local Assembly 2	NB87	CCTGCTCATGGTGTCCTTT	scaffold_2: 6572035-6572053
		NB88	GTCCTACTCCATCAATCCCTTC	scaffold_2: 12563455-12563434
	Large Deletion Local Assembly 3	NB93	GGTTGATTGGATTCTGCAAGT	scaffold_2: 6572109-6572089
		NB94	CTCTGAGTGGAAATAGCATGAATTA	scaffold_2: 12563384-12563408
Translocation TRA1/3	Translocation Local Assembly	NB13	GAGACAAGCACCCCCAATG	scaffold_1: 27525506-27525524
		NB14	TGAAAAGCCAATGATAAGAATTTG	scaffold_3: 31452462-31452485 [z]
	Large Inversion Local Assembly 1	NB143	ACACATCTCCCTCTCTCTCTAC	scaffold_4: 3540337-3540358
		NB144	GCCAATCTAGGAGGGCTATC	scaffold_4: 7515337-7515356 [z]
Large Inversion INV4	Large Inversion Local Assembly 2 Set 1	NB115	CGGGTGTCCAGAGAGAGA	scaffold_4: 3540444-3540427
		NB116	GGTGTACCCTATTCATTAAGGTTG	scaffold_4: 7515430-7515407
	Large Inversion Local Assembly 2 Set 2	NB153	CGGGTGTCCAGAGAGAGA	scaffold_4: 3540444-3540427
		NB154	ACTAGTGACAACCTTCAACGTAA	scaffold_4: 7515539-7515517

[z] primer had multiple hits in BLAST against the reference genome.

Table 4.3 Summary of Real-Time PCR Results for Tango Pollen with Structural Variant Primers

Structural variant	Primer set	Number of samples successfully amplified	Number of samples failed to amplify
TRA1/3	NB13 + NB14	4	3
DEL2	NB81 + NB82	7	0
	NB87 + NB88	6	1
	NB93 + NB94	7	0
INV4	NB143 + NB144	6	1
	NB115 + NB116	6	1
	NB153 + NB154	5	2

Table 4.4 Mean ΔCt Values of Tango Fruit Albedo and Juice Vesicles DNA with Structural Variant Primers

		Mean ΔCt						
	Structural variant primer set	**TRA1/3 NB13+NB14**	**DEL2 NB81+NB82**	**DEL2 NB87+NB88**	**DEL2 NB93+NB94**	**INV4 NB143+NB144**	**INV4 NB115+NB116**	**INV4 NB153+NB154**
Albedo	**Mean**	0.98	1.14	0.56	1.05	0.27	1.19	0.96
	Standard deviation	0.44	0.39	0.20	0.34	0.51	0.31	0.51
Juice vesicles	**Mean**	2.35	2.54	2.35	2.43	1.76	2.86	2.39
	Standard deviation	0.54	0.66	0.60	0.66	0.59	0.59	0.67
Student's t-test t-value		-5.30883	-5.21981	-8.77636	-5.43662	-5.07816	-7.28764	-4.63269
p-value		0.000142	0.000165	0.00001	0.000114	0.000212	0.00001	0.000469
ALB/JV CN*		2.583512	2.642555	3.442206	2.60509	2.798525	3.191718	2.693845

* ALB/JV CN: copy number ratio of chimeric sequence in albedo versus juice vesicle, which is calculated as follow:

$$\frac{albedo\ CN_{SV}}{juice\ vesicle\ CN_{SV}} = 2^{-\Delta\Delta Ct} = 2^{-(\overline{\Delta Ct_{albedo}} - \overline{\Delta Ct_{juice\ vesicle}})}$$

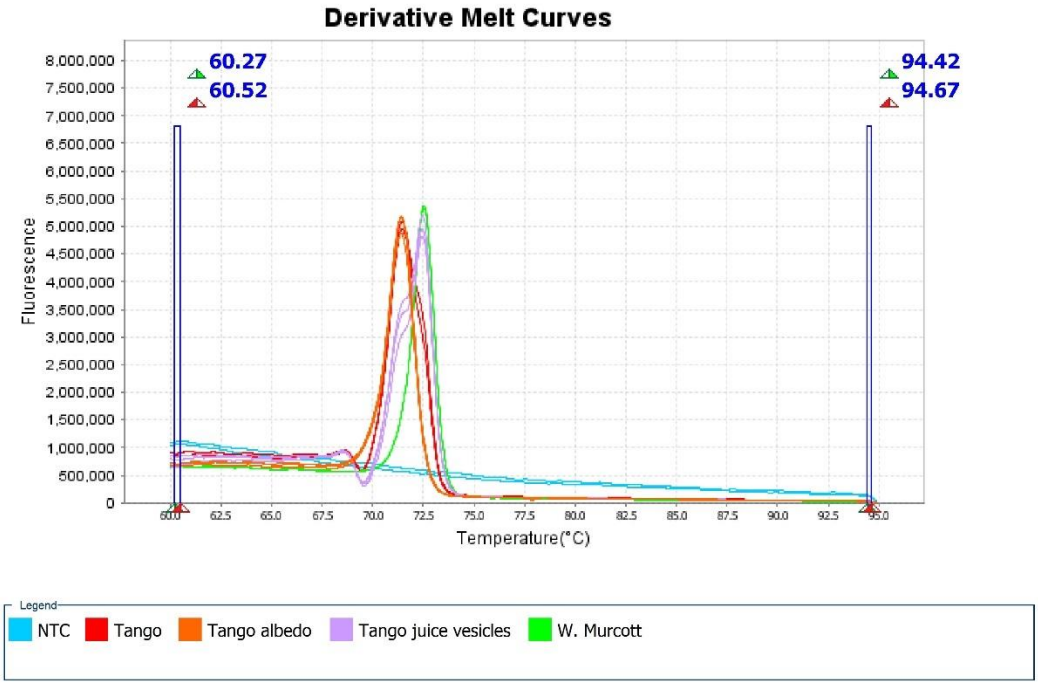


Figure 4.1 Real-Time PCR Derivative Melt Curves of DC-1 Short Deletion with Primer Set YZ_DC1_7F_SYBR and YZ_DC1_7R_SYBR on W. Murcott Leaf Samples, Tango Leaf Samples, Tango Albedo Samples and Tango Juice Vesicle Samples. Red: Tango leaf, orange: Tango albedo, purple: Tango juice vesicle, green: W. Murcott leaf samples with blue: NTC control. Tango leaf samples had a single peak at 71 ℃ while W. Murcott leaf samples had a single peak at 72.5 ℃. Tango albedo samples had a single peak at the same temperature as in Tango leaf samples but the peak is narrower than that of Tango leaf samples. Tango juice vesicle samples had a single peak at the same temperature as in W. Murcott leaf samples with a shoulder at a lower temperature, approximately at 71 ℃. NTC did not amplify.

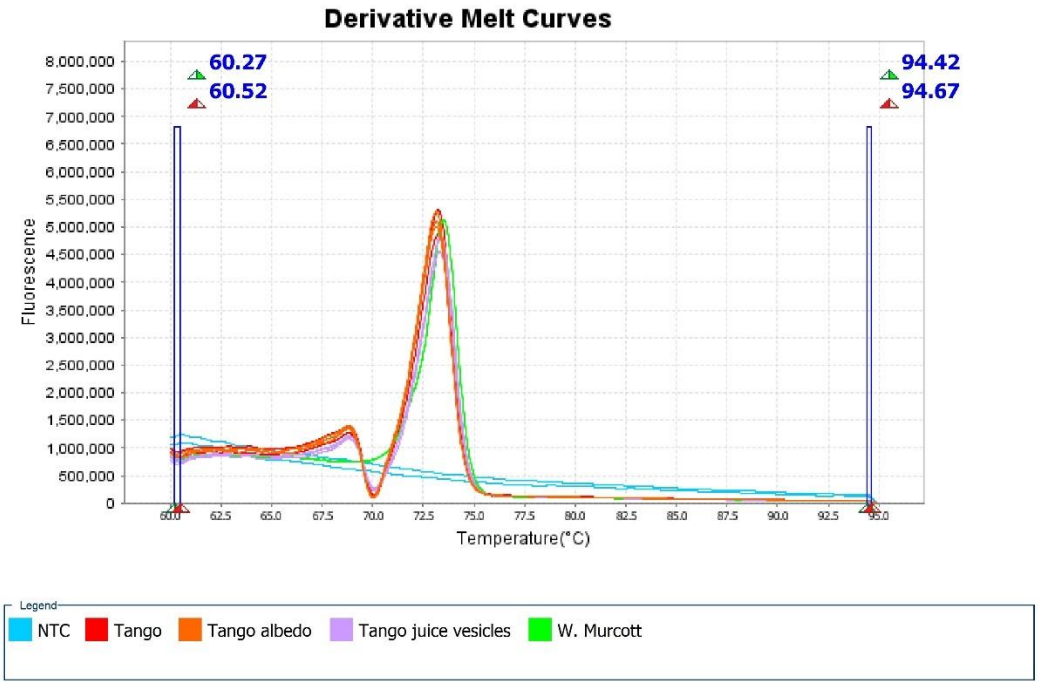


Figure 4.2 Real-Time PCR Derivative Melt Curves of DC-7 Short Deletion with Primer Set UCR56 and UCR57 on W. Murcott Leaf Samples, Tango Leaf Samples, Tango Albedo Samples and Tango Juice Vesicle Samples. Red: Tango leaf, orange: Tango albedo, purple: Tango juice vesicle, green: W. Murcott leaf samples with blue: NTC control. W. Murcott leaf samples had a single peak at 73 ℃, while all Tango leaf, albedo and juice vesicle samples had the same major peak at 73 ℃ with a minor peak at 69 ℃. NTC did not amplify.

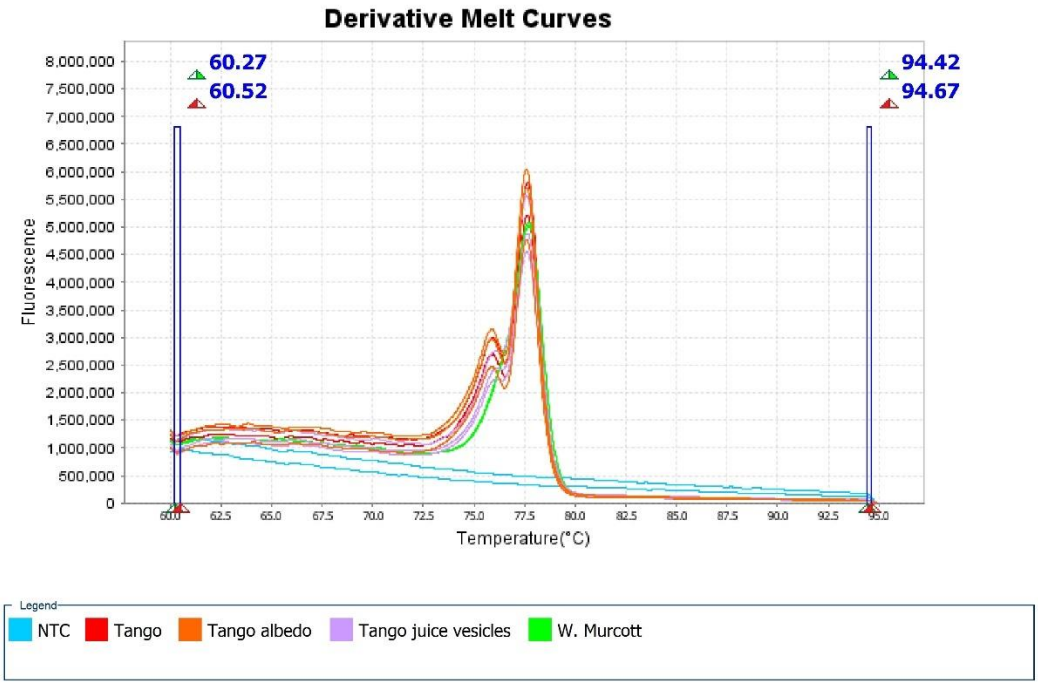


Figure 4.3 Real-Time PCR Derivative Melt Curves of Del2 Short Deletion with Primer Set UCR27 and UCR28 on W. Murcott Leaf Samples, Tango Leaf Samples, Tango Albedo Samples and Tango Juice Vesicle Samples. Red: Tango leaf, orange: Tango albedo, purple: Tango juice vesicle, green: W. Murcott leaf samples with blue: NTC control. W. Murcott leaf samples had a single peak at 77.5 ℃, while all Tango leaf, albedo and juice vesicle samples had the same major peak at 77.5 ℃ with a minor peak at 75.5 ℃. NTC did not amplify.

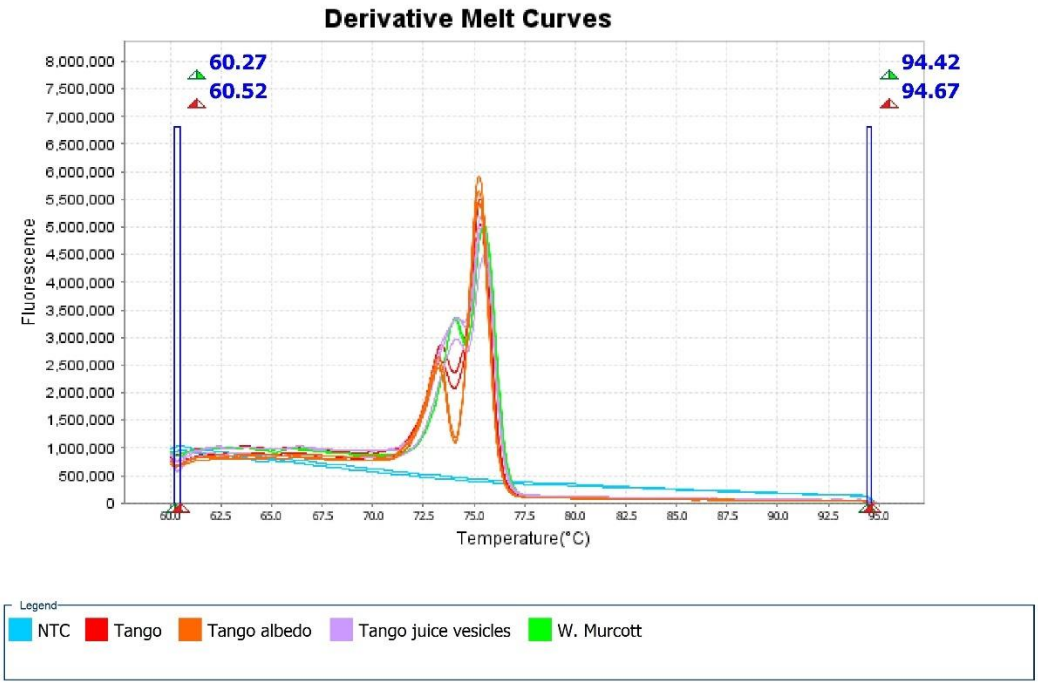


Figure 4.4 Real-Time PCR Derivative Melt Curves of Del3 Short Deletion with Primer Set UCR29 and UCR30 on W. Murcott Leaf Samples, Tango Leaf Samples, Tango Albedo Samples and Tango Juice Vesicle Samples. Red: Tango leaf, orange: Tango albedo, purple: Tango juice vesicle, green: W. Murcott leaf samples with blue: NTC control. W. Murcott leaf samples had a major peak at 75 ℃ with a minor peak at 74 ℃. All Tango leaf, albedo and juice vesicle samples had the same major peak at 75 ℃, while Tango leaf and albedo samples had a minor peak at 73 ℃ and Tango juice vesicle samples had a minor peak at 74 ℃ as in the minor peak in W. Murcott leaf samples. NTC did not amplify.

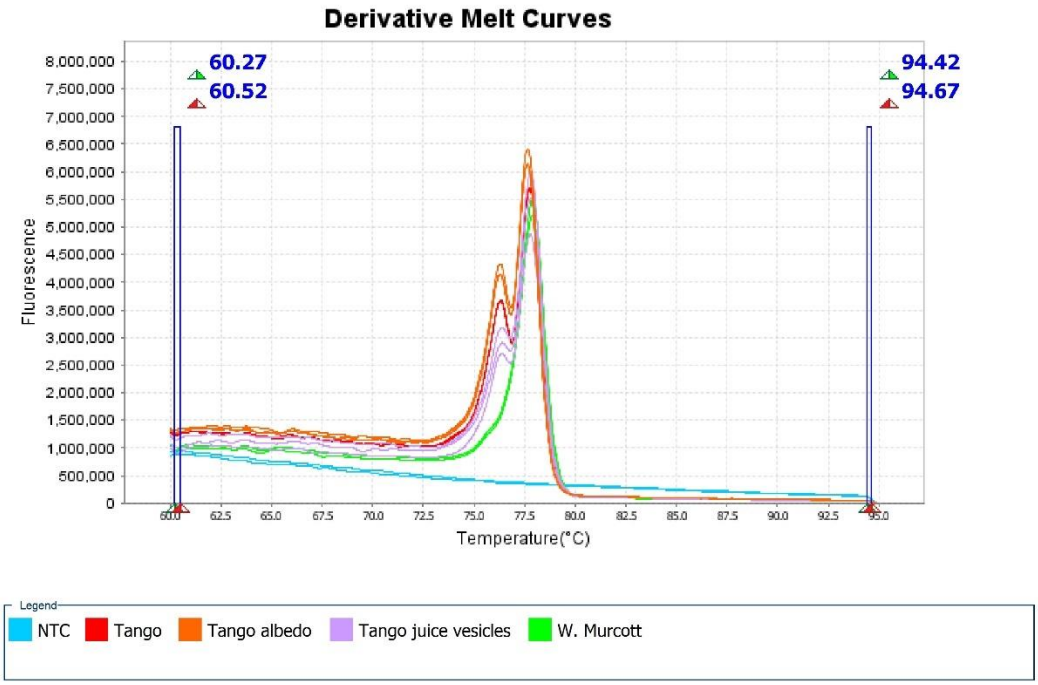


Figure 4.5 Real-Time PCR Derivative Melt Curves of Del4-2 Short Deletion with Primer Set UCR35 and UCR36 on W. Murcott Leaf Samples, Tango Leaf Samples, Tango Albedo Samples and Tango Juice Vesicle Samples. Red: Tango leaf, orange: Tango albedo, purple: Tango juice vesicle, green: W. Murcott leaf samples with blue: NTC control. W. Murcott leaf samples had a single peak at 77.5 ℃, while all Tango leaf, albedo and juice vesicle samples had the same major peak at 77.5 ℃ with a minor peak at 76 ℃. NTC did not amplify.

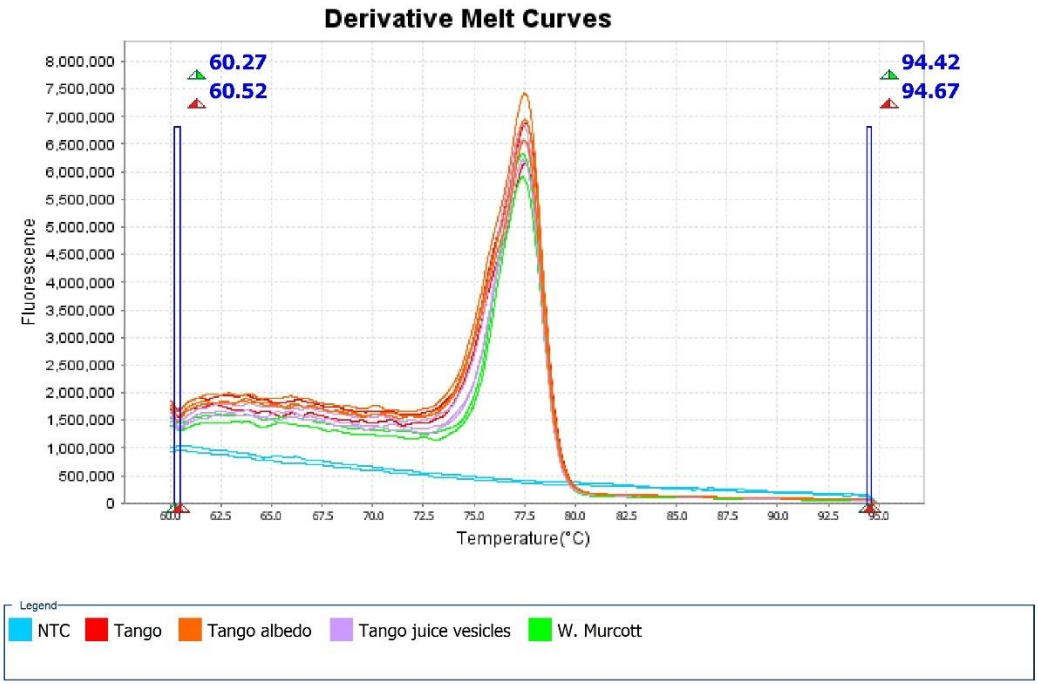


Figure 4.6 Real-Time PCR Derivative Melt Curves of Del3-3 Short Deletion with Primer Set UCR48 and UCR49 on W. Murcott Leaf Samples, Tango Leaf Samples, Tango Albedo Samples and Tango Juice Vesicle Samples. Red: Tango leaf, orange: Tango albedo, purple: Tango juice vesicle, green: W. Murcott leaf samples with blue: NTC control. All W. Murcott leaf, Tango leaf, Tango albedo and Tango juice vesicle samples had a single peak at 77.5 ℃, while all Tango leaf, albedo and juice vesicle samples had slightly broader peak, with more fluorescence at the lower temperature than that of W Murcott leaf samples. NTC did not amplify.

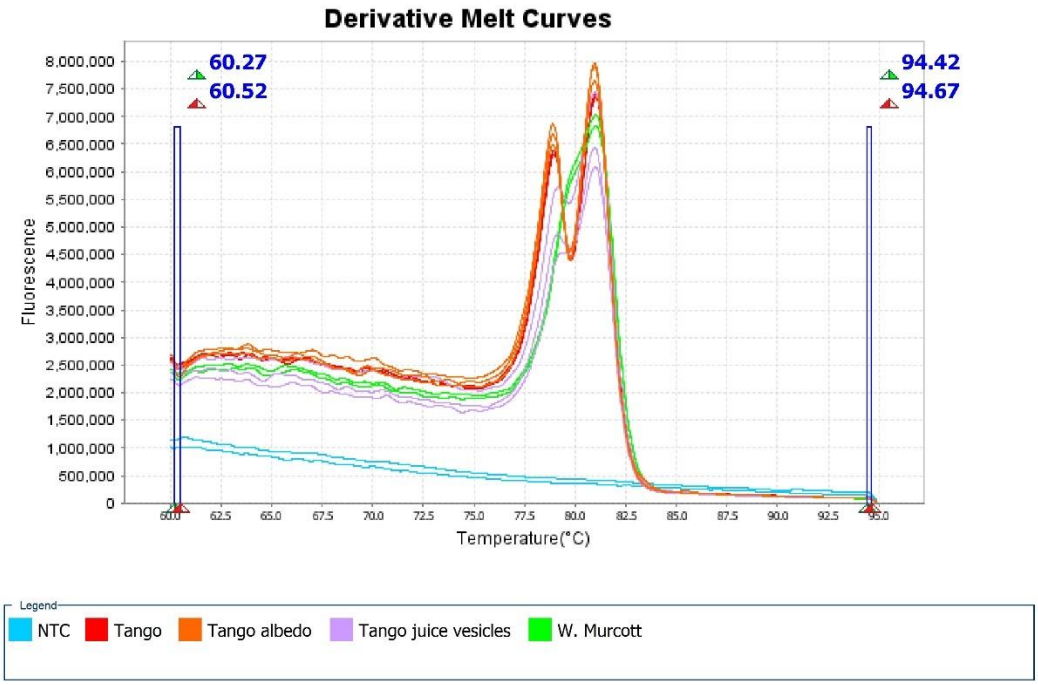


Figure 4.7 Real-Time PCR Derivative Melt Curves of Del2-3 Short Deletion with Primer Set UCR80 and UCR81 on W. Murcott Leaf Samples, Tango Leaf Samples, Tango Albedo Samples and Tango Juice Vesicle Samples. Red: Tango leaf, orange: Tango albedo, purple: Tango juice vesicle, green: W. Murcott leaf samples with blue: NTC control. W. Murcott leaf samples had a single peak at 81 ℃ with a shoulder at the lower temperature side, while all Tango leaf, albedo and juice vesicle samples had a major peak at 81 ℃ with a minor peak at 79 ℃. NTC did not amplify.

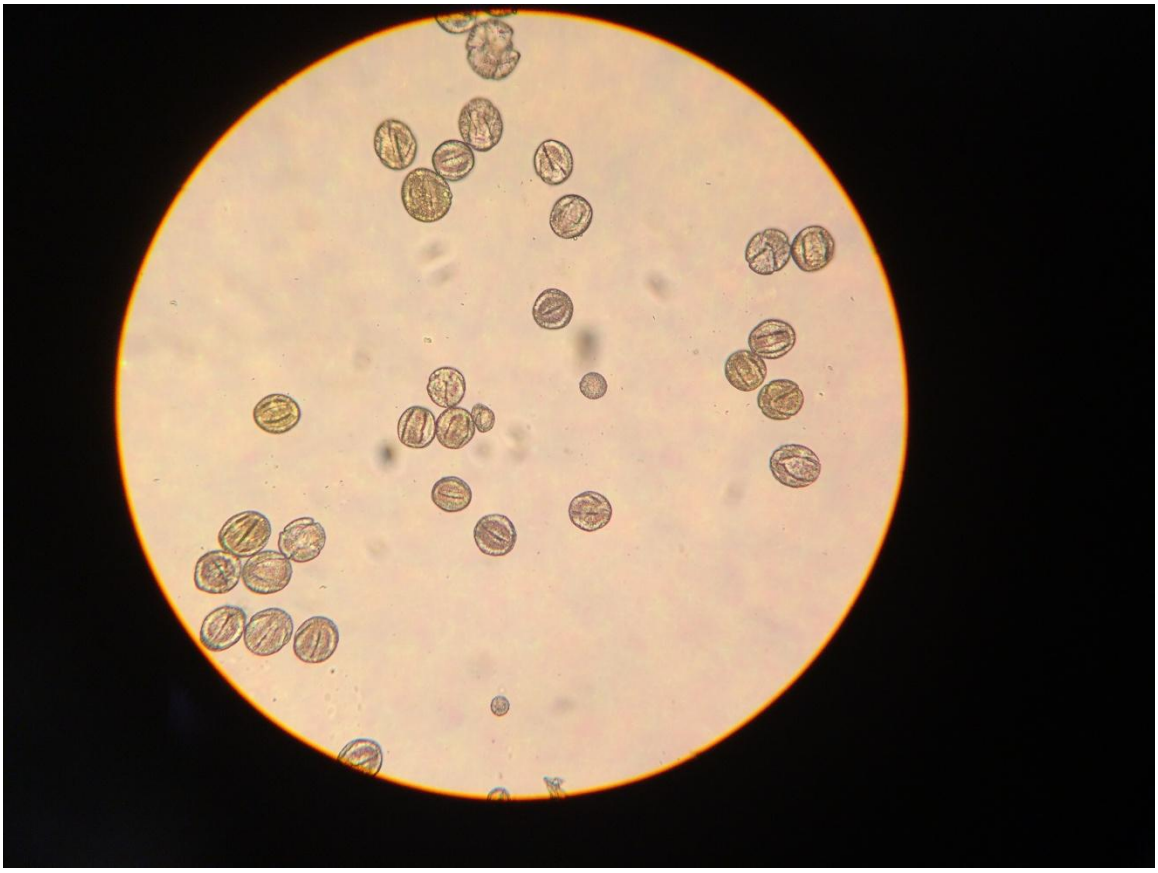


Figure 4.8 Pollen Under the 400X Binocular Microscope After Pollen Isolation. No tissue other than pollen was observed.

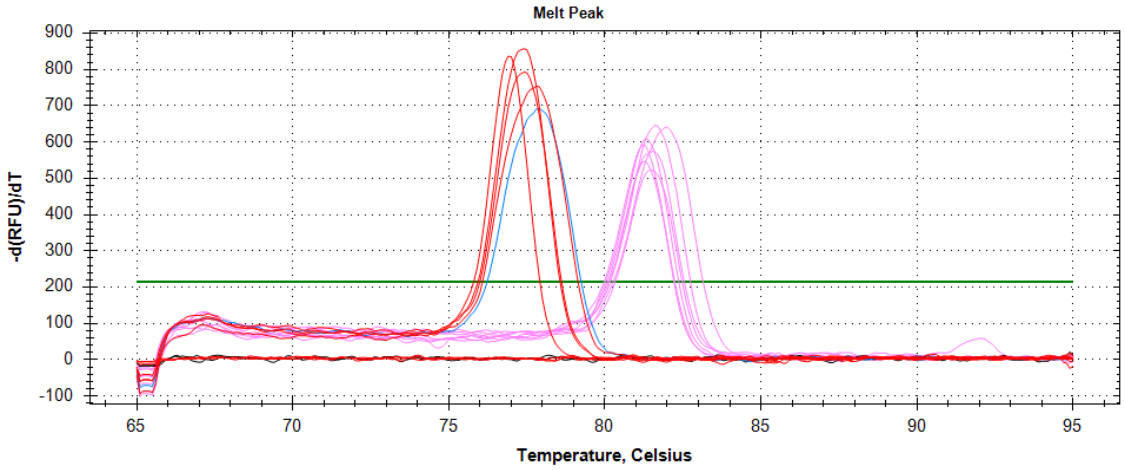


Figure 4.9 Amplification of Seven Tango samples with Translocation Primers NB13 and NB14 which Target the Translocation Involving Chromosome 1 and Chromosome 3. Red: Tango pollen DNA with TRA1/3 primers NB13 and NB14, blue: Tango leaf DNA with TRA1/3 primers NB13 and NB14, black: NTC with TRA1/3 primers NB13 and NB14, pink: Tango pollen DNA with MDH primers Cit1249 and Cit1250, green: NTC with MDH primers Cit1249 and Cit1250.

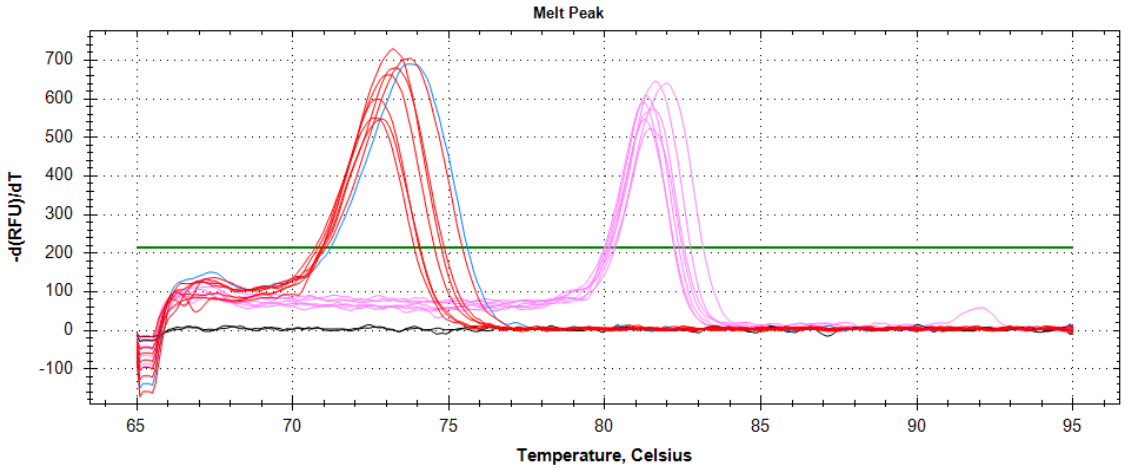


Figure 4.10 Amplification of Seven Tango Samples with Deletion Primers NB81 and NB82 which Target the Deletion on Chromosome 2. Red: Tango pollen DNA with DEL2 primers NB81 and NB82, blue: Tango leaf DNA with DEL2 primers NB81 and NB82, black: NTC with DEL2 primers NB81 and NB82, pink: Tango pollen DNA with MDH primers Cit1249 and Cit1250, green: NTC with MDH primers Cit1249 and Cit1250.

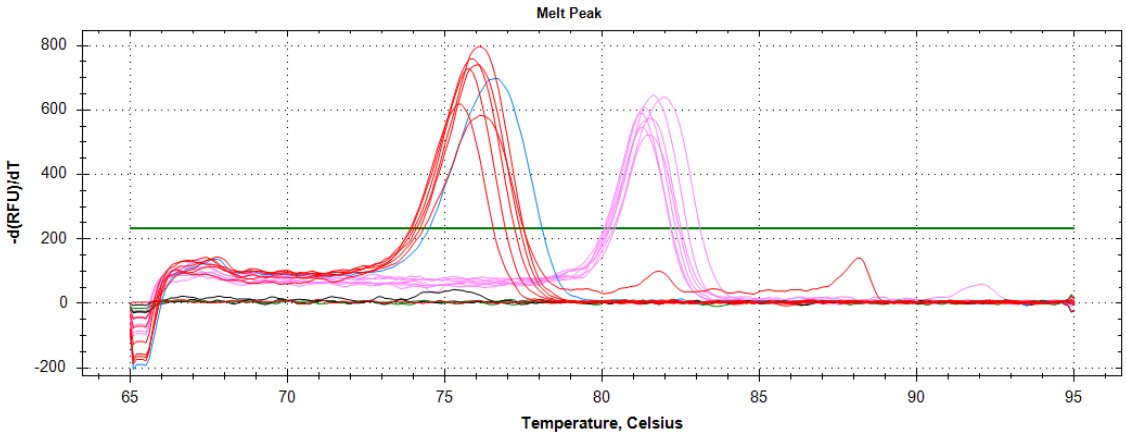


Figure 4.11 Amplification of Seven Tango Samples with Deletion Primers NB87 and NB88 which Target the Deletion on Chromosome 2. Red: Tango pollen DNA with DEL2 primers NB87 and NB88, blue: Tango leaf DNA with DEL2 primers NB87 and NB88, black: NTC with DEL2 primers NB87 and NB88, pink: Tango pollen DNA with MDH primers Cit1249 and Cit1250, green: NTC with MDH primers Cit1249 and Cit1250.

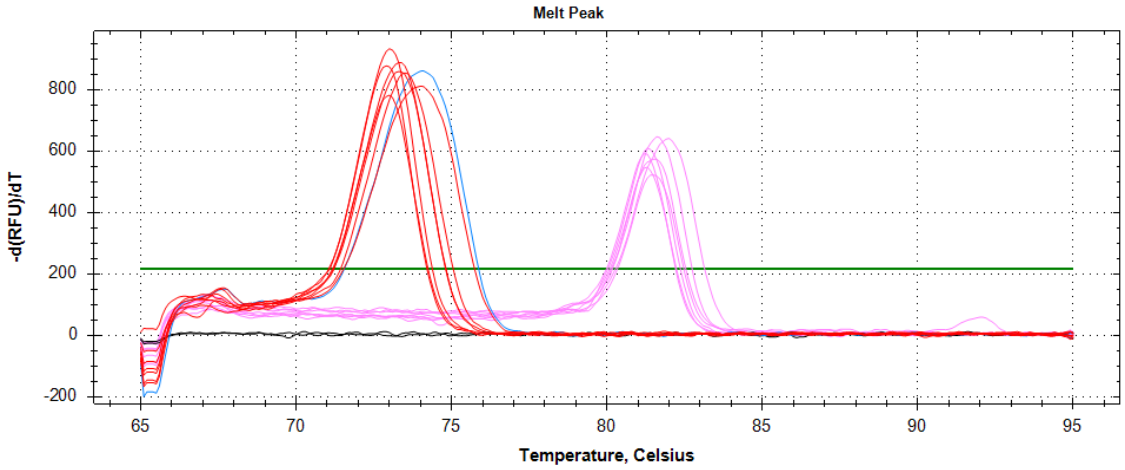


Figure 4.12 Amplification of Seven Tango Samples with Deletion Primers NB93 and NB94 which Target the Deletion on Chromosome 2. Red: Tango pollen DNA with DEL2 primers NB93 and NB94, blue: Tango leaf DNA with DEL2 primers NB93 and NB94, black: NTC with DEL2 primers NB93 and NB94, pink: Tango pollen DNA with MDH primers Cit1249 and Cit1250, green: NTC with MDH primers Cit1249 and Cit1250.

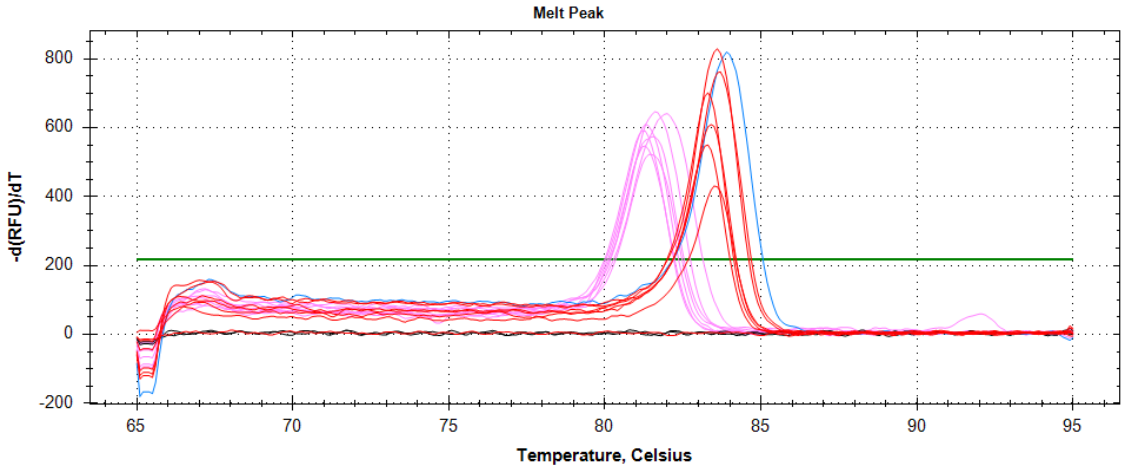


Figure 4.13 Amplification of Seven Tango Samples with Inversion Primers NB143 and NB144 which Target the Inversion on Chromosome 4. Red: Tango pollen DNA with INV4 primers NB143 and NB144, blue: Tango leaf DNA with INV4 primers NB143 and NB144, black: NTC with INV4 primers NB143 and NB144, pink: Tango pollen DNA with MDH primers Cit1249 and Cit1250, green: NTC with MDH primers Cit1249 and Cit1250.

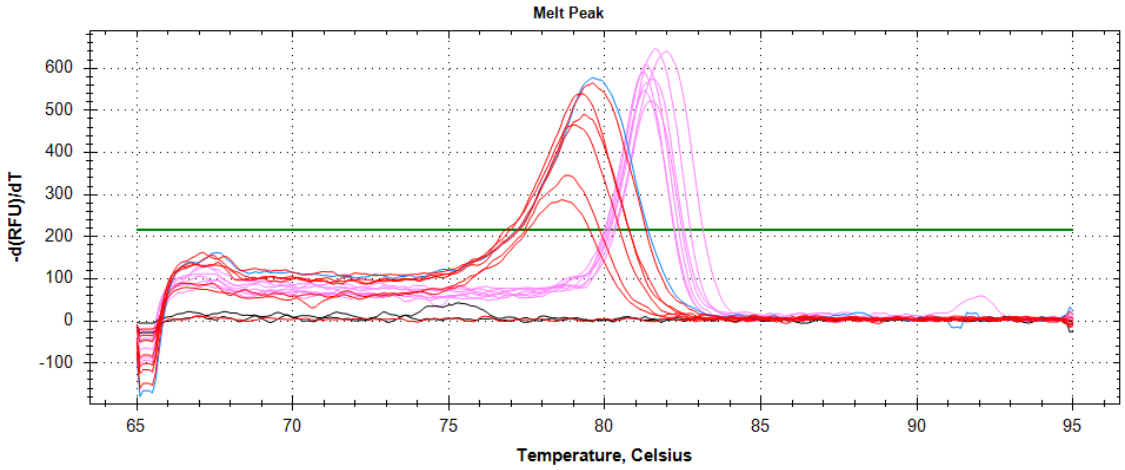


Figure 4.14 Amplification of Seven Tango Samples with Inversion Primers NB115 and NB116 which Target the Inversion on Chromosome 4. Red: Tango pollen DNA with INV4 primers NB115 and NB116, blue: Tango leaf DNA with INV4 primers NB115 and NB116, black: NTC with INV4 primers NB115 and NB116, pink: Tango pollen DNA with MDH primers Cit1249 and Cit1250, green: NTC with MDH primers Cit1249 and Cit1250.

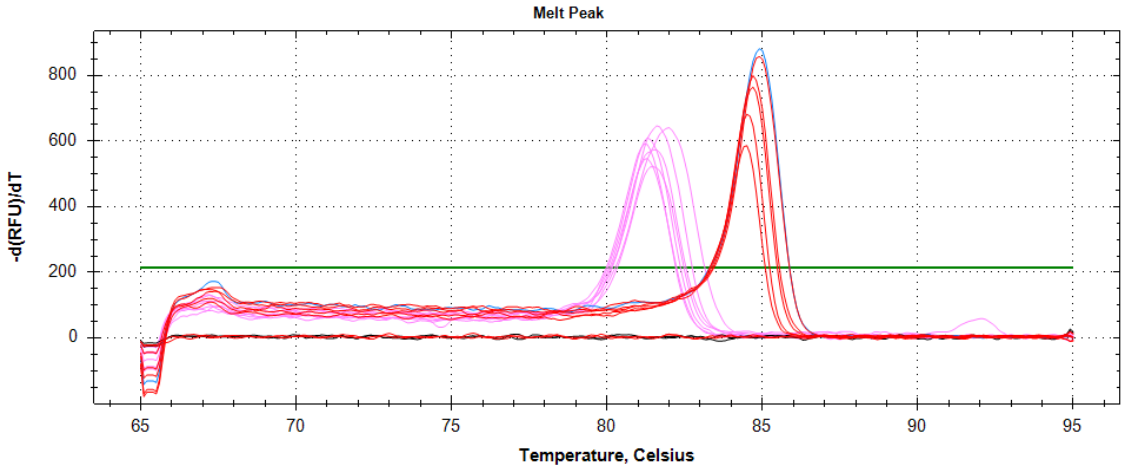


Figure 4.15 Amplification of Seven Tango Samples with Inversion Primers NB153 and NB154 which Target the Inversion on Chromosome 4. Red: Tango pollen DNA with INV4 primers NB153 and NB154, blue: Tango leaf DNA with INV4 primers NB153 and NB154, black: NTC with INV4 primers NB153 and NB154, pink: Tango pollen DNA with MDH primers Cit1249 and Cit1250, green: NTC with MDH primers Cit1249 and Cit1250.

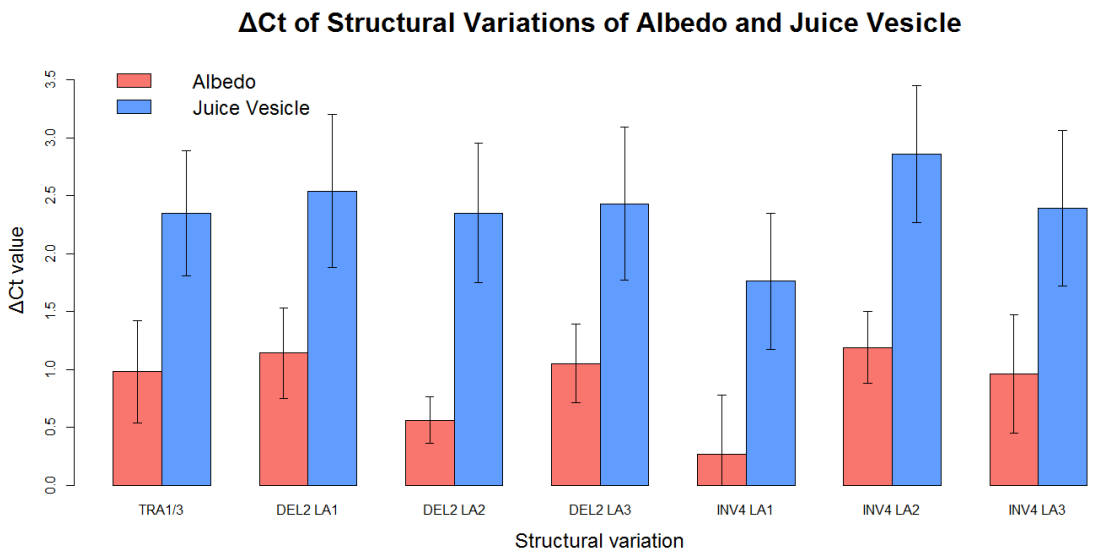


Figure 4.16 Comparison of Mean ΔCt Values of Seven Local Assemblies (LA) of Three Putative Structural Variations (TRA1/3, DEL2 and INV4) Between Albedo DNA (Red Bars) and Juice Vesicle DNA (Blue Bars) of Tango. Primer pairs used for amplification were TRA1/3: NB13 and NB14, DEL2 LA1: NB81 and NB18, DEL2 LA2: NB87 and NB88, DEL2 LA3: NB93 and NB94, INV4 LA1: NB143 and NB144, INV4 LA2: NB115 and NB116, and INV4 LA3: NB153 and NB154. Error bars represent standard deviation.

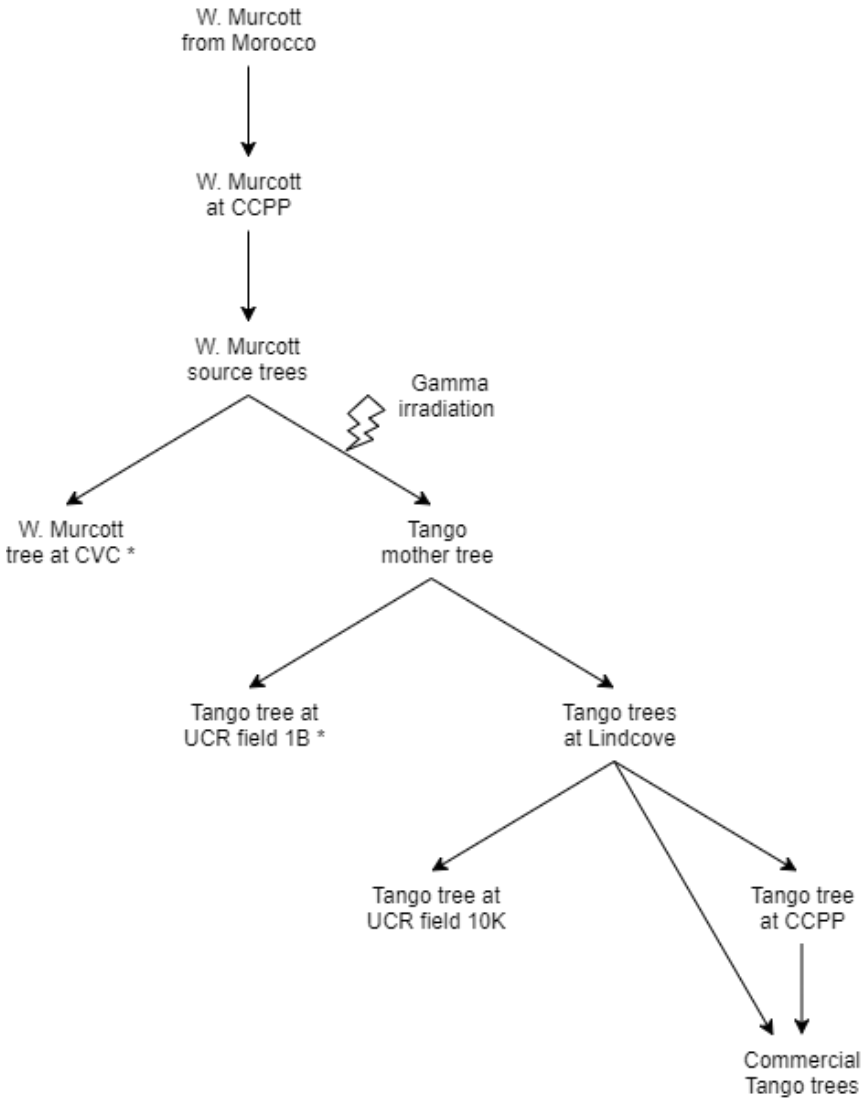


Figure 4.17 The Propagation History of the W. Murcott and Tango Trees Studied. A W. Murcott source tree was propagated by the Citrus Clonal Protection Program (CCPP) using buds obtained from Morocco and planted in 1989. Two additional W. Murcott trees were planted in 1989 in the Citrus Variety Collection (CVC) at University of California, Riverside and leaf tissue from one of them (asterisk) was used for Illumina sequencing. W. Murcott budwood from an unidentified source tree at CCPP was used for irradiation and produced the Tango mother tree in 1995. The clonally propagated descendants of the Tango mother tree were planted at UCR-1B field (asterisk) in 2001, and leaf tissue from one of these trees was used for Illumina sequencing. Additional Tango trees were propagated and planted at UC Lindcove Research and Extension Center. Tango trees at UCR-10K field were propagated from buds of these trees. Leaf tissue, fruit tissue, and pollen used in our studies were collected from Tango trees at UCR-10K field. A new Tango bud source was developed by CCPP and provided additional buds for commercial Tango trees.

Conclusion to Dissertation

Small-Scale Sequence Differences Between W. Murcott and Tango

Two Indels (DC-1 and DC-7) and six SNPs (1-GP, 3-GP, 5-GP, 7-GP, 33-GP and 45-GP) were reported by IVIA in Spain as distinguishing Tango from W. Murcott, however both Indels and only three SNPs (DC-1, DC-7, 3-GP, 5-GP and 33-GP) were verified to show polymorphism between the Tango and W. Murcott trees sampled in California, while the other three SNPs (1-GP, 7-GP and 45-GP) were not observed. In our studies, five additional Indels (Del2, Del3, Del4-2, Del3-3 and Del2-3) were predicted from analysis of Tango and W. Murcott NGS data. These three SNPs and seven Indels altogether were verified on both varieties grown in California by using real-time PCR and conventional PCR followed by sequencing, and each was heterozygous in Tango and homozygous in W. Murcott. None of these SNPs or Indels altered the sequence of genes inferred to function in gamete or seed development. Also, we demonstrated the feasibility of using the discovered SNPs and Indels as molecular markers for discrimination of Tango trees and W. Murcott trees sampled in Riverside, California. Future studies may include testing more Tango and W. Murcott trees grown at other locations and in other countries to determine if these SNPs and Indel markers are characteristic of all Tango trees, and to trace the different propagation generations of Tango to identify whether the SNPs and/or Indels are associated with initial gamma irradiation or developed from natural mutations that occurred during later propagation of Tango.

Chromosome Rearrangements of Tango

In comparison to W. Murcott, Tango has one heterozygous translocation of a 6 Mb segment (from 6572073rd bp to 12563423rd bp) on chromosome 2 (DEL2) and one heterozygous inversion of 4 Mb (from 43540412th bp – 3540416th bp to 7515394th bp – 7515405th bp) on chromosome 4 (INV4) verified from all the chromosome rearrangement predictions. However, the two distal ends of a 6 Mb segment on chromosome 2 were joined and the interstitial segment was apparently translocated from one of the chromosome 2 homologs to an unknown location in the genome. A predicted translocation, TRA1/3 (breakpoints at 27525574th bp of chromosome 1 and 31452544th bp at chromosome 3), was chosen for PCR analysis and it was found rather unlikely to be a translocation. The predicted translocation could be an insertion of a transposon carrying a sequence from at least 31452459th bp to 31452539th bp on chromosome 3 to a new position at approximately 27525524th bp on chromosome 1.

Male-Sterility and Female-Sterility of Tango

Since the inversion of 4 Mb on chromosome 4 (INV4) is sufficiently large, it is possible that chromosome loops form during meiosis. If crossover occurs within the paracentric inversion loop, the dicentric bridge will form and the resulting acentric fragment fails to be included in either of the resulting gametes; thus the lagging chromosome can be observed during the anaphase. For deletion and pericentric inversion, it is unlikely that the homologous chromosomes will lag at the spindle midzone. Either a large translocation or

a large inversion in the heterozygous state could reduce the resulting gametes carrying the whole genome to a half, in which one quarter carry the normal chromosomes and the other quarter carry the inverted chromosome for both paracentric inversion and pericentric inversion. Also, the 6 Mb segment apparently deleted from chromosome 2 may have translocated to another region of chromosome 2 or to a different chromosome, further reducing the proportion of functioning gametes generated. The above results are consistent with previous studies showing the low viability and low germination rate of Tango pollen, as well as lagging chromosome observed at anaphase of meiosis of Tango gametic cells. Therefore, it is likely that the putative translocation on chromosome 2 (DEL2) and the inversion on chromosome 4 (INV4) contribute to the inviability of both Tango male and female gametes and thus result in the inability to fertilize other citrus varieties and seedlessness of Tango fruits. The sequences at breakpoint junctions found in Tango can be used to develop conventional PCR primers flanking the breakpoint, the presence of whose product would be a criterion to simply and rapidly distinguish Tango from W. Murcott.

As noted previously in Chapter IV, the amplification in Tango pollen DNA by structural variation primer sets proves that the previously discovered structural variations, TRA1/3, DEL2 and INV4, are present in cell layer II of the meristem, and could be responsible for the misalignment of chromosomes during meiosis. The subsequent potential crossovers may further result in aneuploid gametes, leading to both male and female gametes developed from these being non-functional so that seedless fruits are thus produced.

Chimerism of Tango

The qualitative analysis of short deletions within fruit albedo and juice vesicle (cell layer I-derived) tissues of Tango suggest that Tango is chimeric with two short deletions, DC-1 and Del3, and three chromosome rearrangements, TRA1/3, DEL2 and INV4, carried by cell layer II as well as its derivatives and likely not in cell layer I. For five short deletions DC-7, Del2, Del4-2, Del3-3 and Del2-3, Tango is uniform in cell layer I and cell layer II of the meristem. The quantitative analysis of copy number of structural variant sequences in comparison to the malate dehydrogenase (MDH) gene within the albedo tissue indicates that the structural variant sequences are present in about half the copy number of MDH gene sequences, supporting the heterozygosity of chromosomal rearrangements of Tango. Also, the significant differences in copy number of structural variant sequences between DNA extracted from albedo and juice vesicles indicated that relatively few cells in juice vesicle tissue carry the chromosomal rearrangement. The marker genotype of cell layer III, which corresponds to the cortex and vascular tissue, was not investigated.

Future Chromosome Analysis

To better characterize the chromosome rearrangements, recent sequencing technologies such as PacBio sequencing and Nanopore sequencing can be applied, which offer longer reads covering more nucleotides surrounding chromosome rearrangement breakpoints and can thus reveal more detailed information about the rearrangement and increase the confidence of chromosome rearrangement prediction. Additional Illumina sequencing

can be used to error correct PacBio or Nanopore reads which tend to have a higher rate of incorrect base calls than Illumina reads.