Characterizing the structural variations in the genome of the mandarin variety, IrM2, induced by gamma irradiation

Upuli Nakandala[1,2], Agnelo Furtado[1,2], Ardashir Kharabian Masouleh[1,2], Malcolm W. Smith[3], Patrick Mason[1,2] and Robert J. Henry[1,2,]* (ID)

[1]*Queensland Alliance for Agriculture and Food Innovation, University of Queensland, Brisbane, Australia*

[2]*ARC Centre of Excellence for Plant Success in Nature and Agriculture, University of Queensland, Brisbane, Australia*

[3]*Department of Primary Industries, Bundaberg Research Station, Bundaberg, Queensland, Australia*

Summary

Fruits with few or no seeds are favoured by consumers because they provide an improved eating experience alongside other important quality traits such as taste and shelf-life. Gamma irradiation has been widely used to induce favourable trait changes in plants, including a reduction in seediness. For this reason, it has been extremely important in the development of new commercial citrus cultivars. The variety IrM2 is a mutant derived from the mandarin variety, Murcott, by gamma irradiation. IrM2 has improved consumer and economic appeal due to its earlier fruit maturity time, low number of seeds and improved external skin colour compared with its progenitor. Here, we developed high-quality, haplotype-resolved genomes for Murcott and IrM2, using PacBio HiFi and Hi-C sequencing. The assemblies ranged from 329 to 344 Mb, with N50s of more than 30 Mb, and more than 98% assembly and annotation completeness for the four haplotypes. Duplications, inversions, translocations and INDELs were the predominant types of mutations found in IrM2. Two large heterozygous inversions (3.1 Mb in Chr3 and 8.6 Mb in Chr6) and one large heterozygous, non-reciprocal translocation (between Chr3 and Chr6) were prominent in IrM2 and may be the causes of the reduced seeds. Variations such as insertions and deletions were also found, resulting in additions and loss of genes in IrM2. The genes lost in IrM2 were associated with many processes, including hormone signalling, flowering, DNA transcription, reproduction, gene expression and transmembrane transport. These high-quality genomes contribute to a deeper understanding of how irradiation affects plant genomes.

Introduction

Mutation breeding using chemical or physical mutagens (radiation) has been widely exploited in the genetic improvements of plants (Mba and Shu, 2012). Mutagenesis has often been adopted in citrus genetic improvements primarily to obtain seedless cultivars but also to improve fruit coloration, disease resistance and induce early ripening (Bermejo *et al*., 2011; Caruso *et al*., 2020). Gamma rays are a source of ionizing radiation, which cause the neutral molecules or atoms to change their natural forms into ionized forms. Gamma rays can dissociate water molecules (radiolysis), producing reactive oxygen species (ROS) that can indirectly induce lipid peroxidation and alterations of the structures of DNA and proteins inside cells. Gamma radiation can also have direct effects on DNA, either by modifying bases of DNA or generating single or double-strand DNA breaks, which can be repaired by natural cell repair mechanisms. Homologous recombination is an error-free repair mechanism while non-homologous end joining is an error-prone repair mechanism, where the latter is more likely to generate mutations such as insertions, deletions and substitutions in the sites of repair. Gamma radiation also can lead to other types of structural aberrations in chromosomes such as inversions and translocations, which can increase the possibility of pollen and ovule abortion (Komura *et al*., 2022; Li *et al*., 2019; Riviello-Flores *et al*., 2022).

The cultivar, Murcott is a *C. reticulata* admixture, which originated in the USA and was later introduced into Australia (Figure S1). Murcott is an important export cultivar but is very seedy (>20 seeds/fruit) and has late season maturity, and so was an obvious candidate for mutation breeding. Consequently, in 1991, Murcott budwood was subjected to gamma irradiation from a cobalt 60 source (^{60}Co) at doses of 30, 40, 50 or 60 Grey, by the Queensland Department of Primary Industries (QDPI), Australia, with the objective of developing low-seeded variants of Murcott. The irradiated buds were budded onto Troyer citrange rootstock and the 136 that survived the treatment were field-planted and maintained until the commencement of fruiting. Fruit from these trees was assessed for seediness, peel colour and maturity time for four consecutive seasons. The variant, IrM2, was a mutation produced at 60 Grey, that not only had greatly reduced seed number (~8 seeds/fruit) but also developed earlier red peel colouration and was edible about two weeks ahead of its progenitor (Smith, 2006) (Figure 1, Figure S1). The fruits from four generations of the vegetatively propagated original IrM2 trees consistently had the same characteristics. IrM2 is in high demand in both local and international markets and has become an important new citrus cultivar. Other low-seeded mutants of Murcott have been produced using the same techniques (e.g. IrM1, Mor, Phoenix: Barry *et al*., 2020; Hardy *et al*., 2017), but IrM2 was the first to have the additional phenotype changes in terms of peel colour
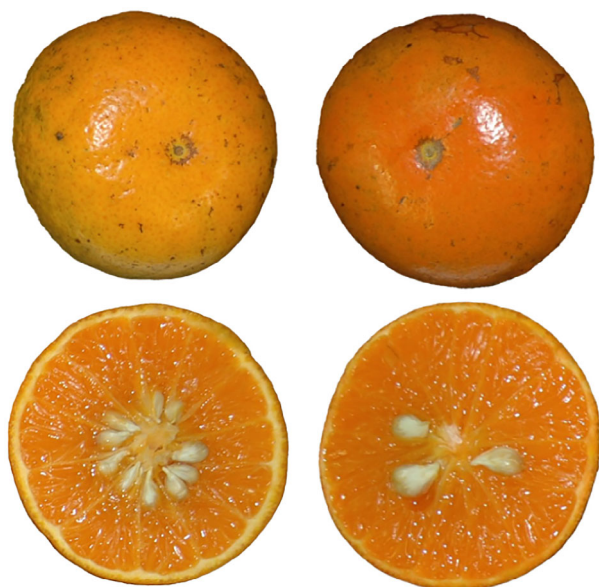
Figure 1 Fruit of Murcott (left) compared with IrM2 (right) shows the stronger peel colour and reduced seed numbers of IrM2 versus its progenitor Murcott.

and maturity time (Caruso et al., 2020) (Barry et al., 2020; Hardy et al., 2017).

The availability of high-quality genomes enables the detection of mutations underlining the desirable phenotypic traits. The present study has developed high-quality, chromosome-level, haplotype-resolved genomes for the varieties Murcott and IrM2, using PacBio HiFi reads and Hi-C reads. The objective of the present study was to identify the chromosomal structural variations between the irradiated genome and its progenitor genome and to explore possible links between these variations and phenotypic traits of interest, particularly those enhancing consumer appeal and economic value.

Results

Genome assembly and annotation of Murcott and IrM2 genomes

The contig assembly and scaffolding of Murcott were performed using 59 Gb of PacBio HiFi reads (~159X from two PacBio SMRT cells) and 115 Gb of Hi-C reads (~351X). Hi-C joined the same contigs to generate most of the scaffolds in the two pipelines that we performed: BWA aligner + Arima mapping and SALSA pipeline and BWA aligner + Arima mapping and YaHS pipeline. However, the joining of the contigs was not correct for some scaffolds in the YaHS pipeline, where it introduced internal telomeres. We compared both pipelines and selected the SALSA scaffolding as the most accurate pipeline in terms of the presence of the telomeres at the ends of the scaffolds. We also re-performed the contig assemblies and the scaffold assemblies with the same pipelines using the same set of input data, and we obtained the same results at the contig and scaffold levels. The scaffold level assembly of Murcott haplotype 1 (MT-hap1) was then aligned with the two haplotype assemblies of *C. sinensis* to assign the scaffolds into pseudochromosomes (Figure S2). *C. sinensis* diploid assemblies were used as reference assemblies to

assign the scaffolds into chromosome level of MT-hap1, due to the fact that *C. sinensis* was one of the parents of Murcott, and a high-quality genome was available for it (The haplotype assemblies of *C. sinensis* were based upon PacBio HiFi, ONT and Hi-C reads). For the MT-hap1, only one scaffold was identified for each of the corresponding chromosomes of *C. sinensis* Haplotype A (Figure S2a) and Haplotype B assemblies (Figure S2b). The Hi-C contact map for the MT-hap1 assembly is shown in Figure S2c. Three pseudochromosomes had telomeres at both ends, while the other six had telomeres at one end (Table S1). Other medium-sized scaffolds ranging from 6 to 1 Mb were included in the final assembly as unplaced/unassigned scaffolds (Nakandala et al., 2024). The sizes of the chromosomes ranged from 25.6 to 50.5 Mb. The total size of the chromosome-level assembly was 303.7 Mb, and the size of the whole genome including the unassigned scaffolds was 335.7 Mb.

The scaffolds of Murcott haplotype 2 assembly (MT-hap2) were aligned against the MT-hap1 pseudochromosomes to assign the scaffolds into pseudochromosomes (Figure S3a). The Hi-C contact map for MT-hap2 assembly is shown in Figure S3b. Three pseudochromosomes had telomeres at both ends, whereas six had telomeres at one end (Table S2). The orientation of the MT-hap2 pseudochromosomes was determined based on the MT-hap1 pseudochromosomes. The size of the pseudochromosomes ranged from 26 Mb to 48.9 Mb, totalling 309.2 Mb, and the size of the whole genome including the unassigned scaffolds was 343.7 Mb.

The contig assembly and scaffolding of IrM2 were performed using 58 Gb of PacBio HiFi reads (~159X from two PacBio SMRT cells) and 86 Gb of Hi-C reads (~263X). The scaffold assemblies of IrM2 haplotypes were independently aligned with Murcott haplotypes. The same set of scaffolds of IrM2 haplotype1 assembly (IrM2-hap1) was aligned with both MT-hap1 and MT-hap2 assemblies, allowing them to be assigned to corresponding pseudochromosomes (Figure S4, Table S3). Four pseudochromosomes had telomeres at both ends, four pseudochromosomes had telomeres at one end and one pseudochromosome had no telomeres at either end. A segment of the scaffold 9 in IrM2 (SC9 1st part) with a telomere at one end was shown to match with Murcott Chr3, while the other part (SC9 2nd part) having a telomere at the other end matched with Murcott Chr6 (Figure S4a,b). Scaffold 11 (SC11), having a telomere at one end, was also shown to match with Chr6 of the Murcott haplotype assemblies (Figure S4). The sizes of SC9 and SC11 were 22 and 6.8 Mb, respectively. The Hi-C contact map for IrM2-hap1 assembly is shown in Figure S4c. The total size of the pseudochromosomes was 306 Mb and that of the whole genome including unplaced scaffolds was 329.3 Mb.

Similarly, the scaffolds of IrM2-hap2 were aligned with both MT-hap1 and MT-hap2 assemblies, and the common scaffolds with high similarity with respect to both assemblies were assigned to corresponding chromosomes (Figure S5a,b). The Hi-C contact map for the IrM2-hap2 assembly is shown in Figure S5c. Out of the nine pseudochromosomes, two had telomeres at both ends and others had a telomere at one end (Table S4). The size of the pseudochromosomes was 309.6 Mb and that of the whole genome including unplaced scaffolds was 336.4 Mb.

The assembly and the annotation statistics of the four whole genome assemblies (nine pseudochromosomes and unplaced scaffolds) are given in Table 1. The whole genome assemblies had more than 98% assembly BUSCO, more than 98.8% annotation

BUSCO and an N50 of more than 30 Mb, revealing the high contiguity and the completeness of the assemblies. The functional annotation of the majority of predicted coding sequences revealed protein descriptions for them. Some sequences which did not have BLAST hits were subjected to coding potential assessment to check the coding potential of them based on both *Citrus* and *Arabidopsis* coding models. The sequences having no BLAST hits in Murcott haplotypes had more than 96% coding potentials (Table S5) and those of IrM2 had more than 98% coding potentials (Table S6).

Comparison between Murcott and IrM2 haplotype assemblies and the determination of the origin of the individual chromosomes of IrM2 haplotypes

Comparison between Murcott haplotype assemblies

The chromosome alignment between the two Murcott haplotypes (Figure S6a,b) revealed large-scale inversions in Chr8 and Chr9, and minor-scale inversions in all the other chromosomes. Other types of sequence variations, including deletions, insertions and structural rearrangements, including translocations, duplications and tandem repeats, were present within all of the chromosomes. The alignment between the whole genomes, including the unplaced scaffolds, also revealed the presence-absence variations (PAV) in each genome (Figure S6c). The results indicated that each assembly had unique sequences (>1 kb blocks) in each chromosome that were not present in the other assembly. Among the unplaced scaffolds, the sequence of SC10 (6.3 Mb) in the MT-hap1 assembly is absent in the MT-hap2 assembly.

Comparison between IrM2 hap1 and Murcott haplotype assemblies

Based on the hypothesis that some chromosomes of IrM2 hap1 might have originated from one Murcott haplotype, while the others might have originated from the second Murcott haplotype, the individual chromosomes of Murcott and IrM2-hap1 assemblies were compared for their structural similarities and differences. Seven chromosomes including Chr1, Chr3, Chr4, Chr5, Chr6, Chr7 and Chr9 of IrM2-hap1 assembly showed high synteny and less structural variation with respect to the corresponding chromosomes of MT-hap1 assembly (Figure 2a, b), indicating that those chromosomes in IrM2-hap1 might have originated from MT-hap1. The PAV analysis also indicated no uniquely present sequences in those seven chromosomes in IrM2-hap1 assembly with respect to MT-hap1 (Figure 2d); however, some sequences were uniquely present in Chr3, Chr5 and unplaced scaffolds in MT-hap1 (Figure 2c).

By contrast, Chr2 and Chr8 of IrM2-hap1 had a high synteny and a low number of structural variations with respect to the MT-hap2 assembly (Figure S7a,b), indicating that these two

chromosomes might have originated from MT-hap2. The PAV analysis also revealed that Chr2 and Chr8 had no uniquely present sequences in IrM2-hap1 with respect to MT-hap2 (Figure S7d) and vice versa (Figure S7c), further confirming the possible origin of these two chromosomes from MT-hap2. Additionally, the sequences related to scaffold 16 (SC16) and scaffold 17 (SC17) were only present in IrM2-hap1 with respect to MT-hap2 (Figure S7d); however, they were not unique in IrM2-hap1 with respect to MT-hap1 (Figure S7d), indicating that SC16 and SC17 might have originated from MT-hap1.

Comparison between IrM2 hap2 and Murcott haplotype assemblies

Similar to the previous hypothesis, some chromosomes of IrM2-hap2 might have derived from one Murcott haplotype, while the others might have derived from the second Murcott haplotype. To investigate that possibility, whole genome alignments were made between the two Murcott assemblies and the IrM2-hap2 assembly. The alignment between MT-hap2 and IrM2-hap2 assemblies revealed that the seven chromosomes (Chr1, Chr3, Chr4, Chr5, Chr6, Chr7 and Chr9) had high synteny with MT-hap2 pseudochromosomes (Figure 3a,b). This suggests that these seven chromosomes of the IrM2-hap2 assembly were derived from MT-hap2. The PAV analysis confirmed the possible derivation of these seven pseudochromosomes in the IrM2-hap2 assembly from MT-hap2, although there were some unique sequences in Chr7 (Figure 3c) and Chr3 (Figure 3d) of Murcott and IrM2, respectively.

By contrast, Chr2 and Chr8 in the IrM2-hap2 assembly had high synteny and fewer structural variations with respect to MT-hap1 (Figure S8a,b), indicating their possible origin from MT-hap1. The likely origin of these two chromosomes was further confirmed by the PAV analysis, where those two chromosomes had no present and absent variations in MT hap1 and IrM2 hap2 assemblies with respect to each other (Figure S8c,d).
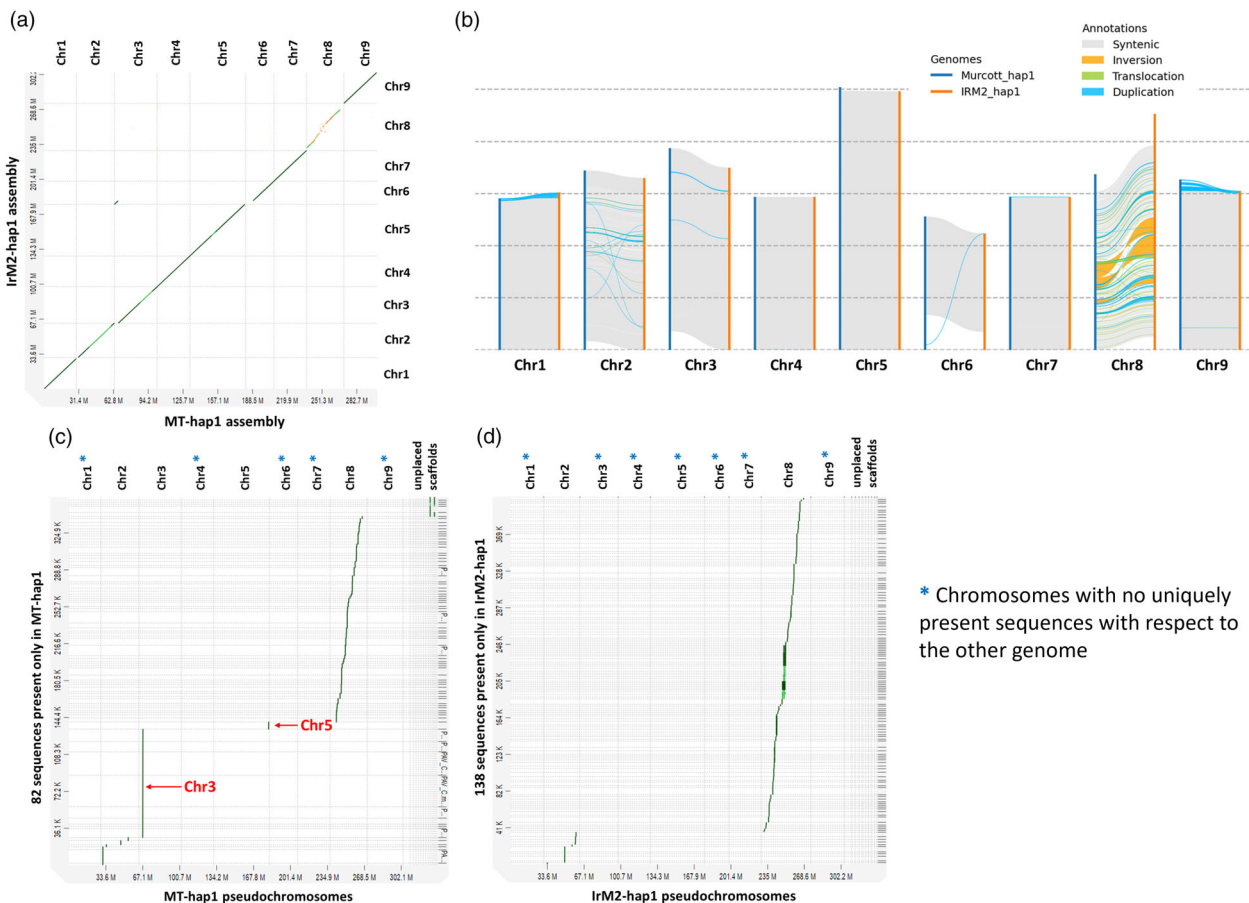
Based on the above structural similarities and differences between the assemblies, the origins of each pseudochromosome of the IrM2 haplotypes from the Murcott assemblies were determined as depicted in Figure 4.

The structural differences in IrM2 haplotypes resulted by irradiation and the associated functional consequences

The whole genome alignment between IrM2-hap1 and Murcott assemblies revealed a non-reciprocal chromosomal translocation in the IrM2-hap1 assembly, where a segment of Chr3 (3.5 Mb) of IrM2-hap1 was detached and translocated to Chr6 of the IrM2-hap1 assembly (Figure S4a,b). The SC9 represented a full chromosome with telomeres at both ends. The SC11 (6.8 Mb) was also a separate part of the Chr6. Therefore, the rearranged Chr6 (SC9) is hereafter denoted as Chr6a (22.4 Mb), and SC11 is denoted as Chr6b (6.8 Mb). The IrM2 assemblies were found to

Table 1 The assembly and annotation statistics of the whole genome assemblies of Murcott and IrM2 varieties

Name of the Variety	Type of the assembly	Size (Mb)	Assembly BUSCO (%)	Assembly N50 (Mb)	No. of genes	No. of CDS	Annotation BUSCO (%)
Murcott	hap1	335.7	98.6	32.7	41 138	45 572	98.8
	hap2	343.7	99.0	32.4	28 715	33 563	99.3
IrM2	hap1	329.3	98.4	30.5	32 669	37 321	99.1
	hap2	336.4	98.8	33.8	37 068	41 530	99.5

Figure 2 The structural similarities and differences between MT-hap1 and IRM2-hap1 assemblies. (a) A part of Chr3 in IrM2-hap1 was translocated to Chr6. (b) Chr1, Chr3, Chr4, Chr5, Chr6, Chr7 and Chr9 of IrM2-hap1 assembly showed high syntenic regions with the corresponding chromosomes of MT-hap1 assembly. Chr2 and Chr8 of IrM2-hap1 assembly have fewer syntenic regions with those of MT-hap1 assembly. (c) Some sequences (>1 kb blocks) were only present in Chr2, Chr3, Chr5, Chr8 and unplaced scaffolds of MT-hap1 assembly, which were absent in IrM2-hap1 assembly. (d) The presence-absence variations between the two assemblies revealed no unique sequences in Chr1, Chr3, Chr4, Chr5, Chr6, Chr7 and Chr9 in IrM2 assembly with respect to MT-hap1 assembly. Some sequences were only present in Chr2 and Chr8 in IrM2-hap1 assembly, which were not present in the corresponding chromosomes of MT-hap1 assembly.

be heterozygous for the translocation, where the translocation could be only observed in the hap1 assembly.

The alignment between the two IrM2 assemblies showed major structural rearrangements, some of which were the result of variations between Murcott haplotypes, whereas the other variations were likely the result of the irradiation treatment (Figure S9). Thus, the genomic rearrangements caused by the irradiation treatment could be differentiated from existing Murcott haplotype variation. The genomic rearrangements present in IrM2 haplotypes with respect to the two Murcott haplotypes were likely the mutations caused by irradiation. By contrast, the rearrangements present in IrM2 haplotypes with respect to one of the Murcott haplotypes were likely due to the variations in between Murcott haplotypes. The dotplots revealed four major structural rearrangements, of which three involved Chr3 and Chr6, and they are likely to be the result of irradiation. The translocation between Chr3 and Chr6 (Figure S9a) may be due to the irradiation treatment and occurred in IrM2-hap1 (Figure S9b). The two inversions in Chr3 (3.1 Mb) and Chr6 (8.6 Mb) (Figure S9a) are also likely be due to the irradiation

treatment and occurred in IrM2-hap2 (Figure S9b). These three structural variations attributed to irradiation are heterozygous for the mutations. By contrast, the inversion found in Chr9 was due to the variations between the Murcott haplotypes (Figure S9).

To investigate the possible consequences of the chromosomal translocation, the translocated segment and the remaining segment of Chr3 were independently aligned with MT-hap1 Chr3 (reference) using the default manual parameters of the long-read support mapping tool. The mapping of these two segments is shown in Figure S10. The scaffolds of IrM2 that were involved in the translocation are shown in Figure S10a. Of the translocated segment, only 2.1 Mb of sequence could be aligned with the reference with several INDELs and many SNPs (Figure S10b,c). A telomeric sequence of 184 bp preceding the aligned region, and a 1.35 Mb sequence following the aligned region of the translocated segment could not be aligned with the reference (Figure S10c). The unaligned region of 1.35 Mb of the translocation segment also did not match with Chr6 of Murcott. In addition, the tool was unable to make an alignment for a 14 kb sequence within the translocated segment of IrM2
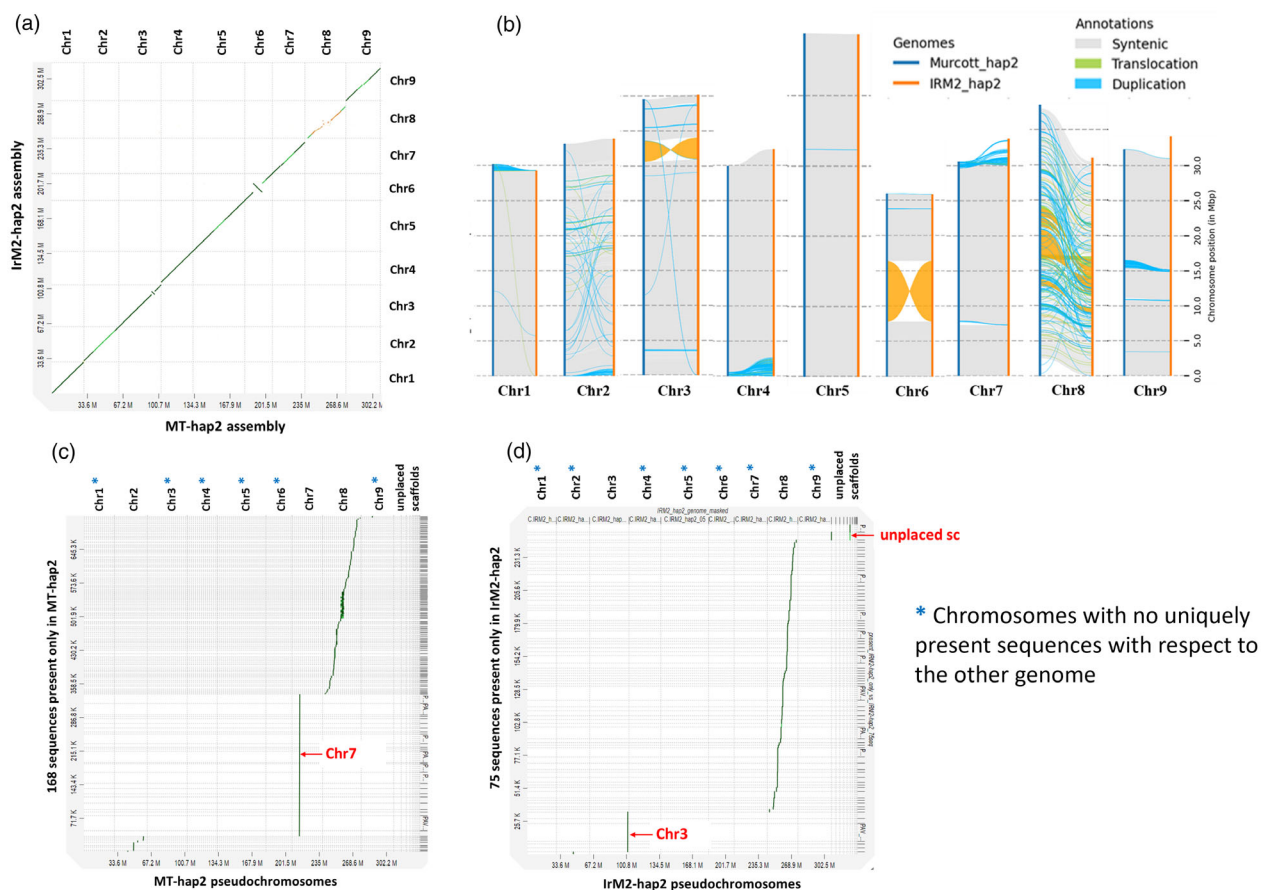
Figure 3 The structural similarities and differences between MT-hap2 and IrM2-hap2 assemblies. (a) The whole genome alignment between the two assemblies. Inversions were found in Chr3 and Chr6. (b) The genome alignment showed that Chr1, Chr3, Chr4, Chr5, Chr6, Chr7 and Chr9 in the two genomes were in high synteny, whereas Chr2 and Chr8 have more structural variations. (c) PAV analysis indicated that Chr2, Chr7 and Chr8 had some uniquely present sequences in the MT-hap2 assembly that are absent in the IrM2-hap2 assembly. (d) Chr2, Chr3, Chr8 and unplaced scaffolds had some uniquely present sequences in the IrM2-hap2 assembly that were absent in the MT-hap2 assembly. All the other chromosomes had no unique sequences in the IrM2-hap2 assembly.
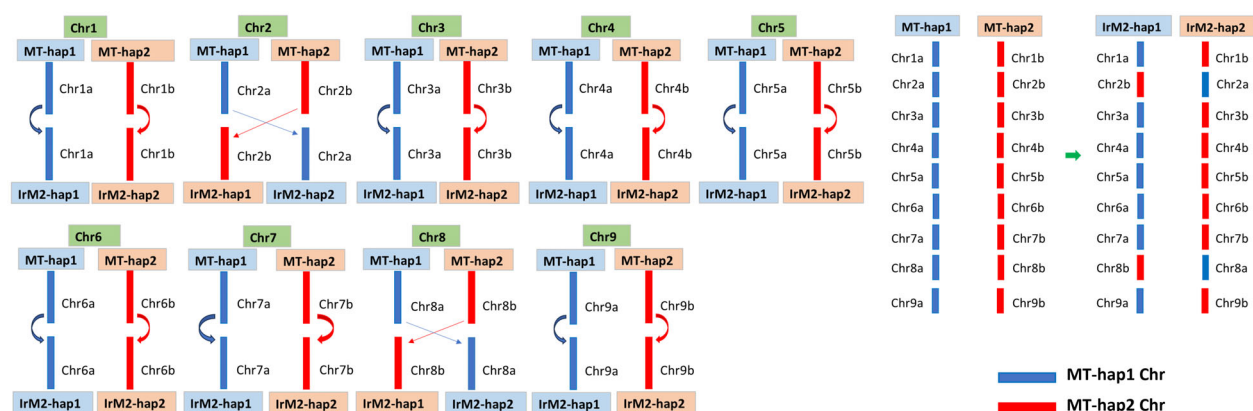
Figure 4 The derivation of chromosomes in IrM2 haplotypes from Murcott haplotypes. The Chr1, Chr3, Chr4, Chr5, Chr6, Chr7 and Chr9 of IrM2-hap1 assembly were derived from MT-hap1 assembly. The Chr2 and Chr8 of IRM2-hap1 were derived from MT-hap2 assembly. The Chr1, Chr3, Chr4, Chr5, Chr6, Chr7 and Chr9 of IrM2-hap2 assembly were derived from MT-hap2 assembly. The Chr2 and Chr8 of IrM2-hap2 were derived from MT-hap1 assembly.

with the reference sequence. The alignment of the start region of the remaining segment of Chr3 of IrM2 with MT-hap1 is shown in Figure S10d. The positions of the end region of the translocated

segment in the alignment (Figure S10c) and the starting region of the remaining aligned segment of the IrM2 Chr3 (Figure S10d) had an intervening gap of 0.163 Mb of sequence in between

them (Figure S10b). This gap indicates that there was no sequence in IrM2-hap1 to cover that region in MT-hap1. The sequences within this region were also found to be uniquely present in MT-hap1 (Figure 2c) and were absent in IrM2-hap1 (Figure 2d) by the PAV analysis. Several genes annotated within this region in MT-hap1 were missing in IrM2-hap1 (Figure S11, Table S7).

Similar to the above alignments, Chr6 of IrM2-hap1 was mapped to Chr6 of the MT-hap1, to explore whether any genes were interrupted due to the detachment of SC11 from Chr6 in IrM2. A gene in Murcott (g18309) was found to overlap at the region where the translocated segment of Chr3 was connected to Chr6 of IrM2-hap1 (Figure S12). This gene g18309 encodes a metal tolerance protein C2. Another gene (g17165) encoding a hypothetical protein in MT-hap1 also was found to be within a Murcott-specific region on Chr5 and was associated with chromosome organization, cellular metabolic process and leaf morphogenesis. Furthermore, the MT-hap2-specific regions in Chr7 were found to have many genes within those regions that were absent in IrM2-hap2 (Table S8). The genes only in the Murcott-specific regions (deleted in IrM2) were found to be involved in many processes including amino acids metabolism, flowering, chromosome organization, regulation of developmental process, protein binding, pollen development, gene expression, oxidoreductase activity, plant hormone signalling (auxin, gibberellin, abscisic acid, ethylene, brassinosteroid and jasmonic acid), chromatin remodelling, plant signal transduction, transmembrane transport, DNA-templated transcription and protein phosphorylation (Table S7, Table S8). The genes that were only present in Chr3 of IrM2-hap2 were found to encode uncharacterized and hypothetical proteins that are associated with protein kinase activities. The missing and inserted genes in IrM2 might be associated with its altered phenotypic traits compared with Murcott mandarin, although specific functional analysis would be required to establish a causal relationship with these particular genes.

The Murcott chromosomes and the IrM2 chromosomes originating from the corresponding Murcott haplotypes were also analysed for their structural differences (Table 2). The main types of structural rearrangements in IrM2-hap1 were the duplications (copy gains or copy losses), inverted duplicates and small intra-chromosomal translocations. The duplications were present in all the pseudochromosomes. There were only copy

Table 2 The structural variations in IrM2 haplotypes compared with Murcott haplotypes in individual chromosomes

Chromosome	IrM2-hap1 variations	IrM2-hap2 variations
Chr1	DUP/INVDP	DUP/INV/INVDP/TRANS/INVTR
Chr2	DUP/INVDP	DUP/INVDP
Chr3	DUP/INVDP/DEL	DUP/INVDP/INS/DEL/INV
Chr4	DUP	DUP/INVDP/TRANS
Chr5	DUP/INVDP	DUP
Chr6	DUP/INV/INVDP/TRANS	DUP/INV
Chr7	DUP	DUP/INVDP/TRANS/DEL
Chr8	DUP/INVDP/TRANS	DUP
Chr9	DUP	DUP/INVDP/TRANS/INVTR

Abbreviations: DEL, deletions (absent sequences); DUP, duplications; INS, insertions (present sequences); INVDP, inverted duplicates; INV, inversions; INVTR, inverted translocations; TRANS, translocations.

gains in Chr1, while there were only copy losses in Chr3, Chr4, Chr5, Chr7 and Chr9. In Chr2, Chr6 and Chr8, both copy gains and losses were found. Of the duplicated regions, there were 1392 inverted duplicated regions, 596 duplicated regions and 245 copy losses in hap1 genome (Table S9). Two intra-chromosomal translocations were found within Chr6 (0.01 Mb) and Chr8 (0.37 Mb). A few smaller inversions, ranging from 178 to 926 bp, were also found in Chr6. In addition, some deletions were found in Chr3 and Chr5 of IrM2-hap1 according to the PAV analysis (Table 2). Among the structural variants, only a few including translocations, inverted translocations and inverted duplications that might have clear breakpoints might be more directly associated with gene fusions. In IrM2-hap1, there was a translocated region within the Chr8 with respect to Murcott Chr8 and was associated with a single gene. The inverted duplicates found in Chr1, Chr3 and Chr8 were associated with a single gene or multiple genes (Table S10).

Duplications, inverted duplicates, intra-chromosomal translocations and inversions were the predominant types of structural variations in IrM2-hap2 (Table 2). Duplications were present in all the pseudochromosomes (1343 duplicates and 150 inverted duplicates) (Table S9). Only copy losses were found in Chr1 and Chr5, while there were both copy gains and losses in other chromosomes (a total of 568 copy losses were identified in hap2, Table S9). Small-scale translocations were found in Chr1 (0.01 Mb – 0.08 Mb), Chr4 (0.001 Mb – 0.05 Mb), Chr7 (0.06 Mb) and Chr9 (0.05 Mb). A small-scale inversion was found in Chr1 (163 bp), and two other large inversions were found in Chr3 and Chr6 as previously mentioned. Additionally, the PAV analysis revealed some insertions in Chr3 and deletions in Chr7 of IrM2-hap2 (Table 2). In IrM2-hap2, the regions with inverted duplicates and translocations in Chr1, Chr2, Chr4, Chr7 and Chr9 were associated with a single or multiple genes that might indicate possible gene fusions.

Discussion

Many studies have explored irradiation-induced mutations in plant genomes using whole genome resequencing, genetic mapping (Hase et al., 2023) and whole genome painting methods (Haskins and Kato, 2019), including studies comparing natural and induce mutations in citrus (Caruso et al., 2019). Our work has expanded on these studies by providing the characterization of irradiation-induced structural variations using high-quality chromosome scale, haplotype-resolved genome assemblies and annotations. Many structural variations were detected, including local sequence variations (SNPs and INDELs), and other complex chromosomal rearrangements such as inversions, translocations and duplications in the mutated IrM2 genome. The regions in the IrM2 genome with translocations, inverted translocations and inverted duplicates (primarily involve breakpoints) that were associated with single or multiple genes were identified, and they might be associated with gene fusions. Importantly, we were able to relate these structural variations in each haplotype of IrM2 with their corresponding chromosomes in the progenitor Murcott haplotypes.

Ionizing radiation is known to cause DNA double-strand breaks in cells which induce chromosomal translocations and genomic instability if they are mis-repaired. Non-homologous end joining is the dominant type of DNA repair mechanism for double-strand breaks caused by ionizing radiation and is more likely to introduce errors during the processing of the ends of the breaks before

ligation (Mahaney *et al.*, 2009). In addition to translocations, double-strand breaks induced by gamma irradiation can also lead to other types of rearrangements including deletions and telomere addition (Myung and Kolodner, 2003). The upregulation of the telomerase and the addition of telomeric repeats are known to protect the broken chromosomal ends resulting from irradiation (Bosco and Haber, 1998; Leteurtre *et al.*, 1997). Therefore, it is possible that the addition of the telomeric repeats at one end of the translocation segment of Chr3 in IrM2 is a result of DNA repair after the irradiation treatment.

Translocations can interrupt or misregulate the regular functions of genes (Farré *et al.*, 2012). The identified translocation from Chr3 to Chr6 was a non-reciprocal, heterozygous translocation. Reciprocal translocations are one type of translocation, that is induced by gamma irradiation (Verma *et al.*, 2019), where two DNA segments are exchanged between two non-homologous chromosomes. Interstitial translocations, on the contrary, involve DNA fragments being inserted within the other chromosomes, instead of the ends of the chromosomes being involved in the translocation, and are also known to result from ionizing radiation (Matsuda *et al.*, 2023). The translocation in IrM2 was only one-way, where the two broken pieces from Chr3 (SC9 1st part) and Chr6 (SC11) were not reciprocally exchanged; instead, SC11 was kept separate without being transferred to Chr3. Many insertions, deletions and SNPs were found within the translocated segment of IrM2, supporting the error-prone nature of the non-homologous end-joining repair mechanism for double-strand breaks.

Previous studies have reported that translocations in plants can cause defects in meiosis and lead to pollen semi-sterility (Tian *et al.*, 2021; Verma *et al.*, 2019). Translocation heterozygotes can lead to meiotic non-disjunction, which leads to abnormal chromosome pairing during meiosis. The fertilization of such a genetically unbalanced gamete with a normal gamete can result in non-viable embryos, which often abort (Forejt, 2001). Different types of chromosomal inversions are also known to have different impacts on reproductive potential. Some inversions can lead to abortions or failures of embryo development, while other inversions have no impact on reproductive potential (Mennuti, 2019). Irradiation-derived mutations of the seedy, self-incompatible and parthenocarpic cultivars such as Moncada mandarin and W. Murcott mandarin have been shown to produce lower numbers of seeds, even when cross-pollinated. The reduced number of seeds in these mutated cultivars is known to be associated with low pollen and ovule viability, leading to male and female sterility (Bermejo *et al.*, 2011; Zhu, 2018). The sterility of the gametes of the mutant of W. Murcott mandarin (Tango) is thought to be due to heterozygous translocations, inversions and deletions in the mutated genome (Zhu, 2018). Therefore, it is possible that the reduced seed numbers seen in IrM2 may also be due to heterozygous translocations and inversions.

In addition to the reduced number of seeds, IrM2 is favoured for its early fruit maturity and improved external peel colour. The genes associated with major deletions in Chr3, Chr5 and Chr7 of IrM2 were associated with many biological processes and molecular functions, including genes related to reproduction, hormone signalling, chromosome organization and development processes. The maturation of citrus fruits is known to be associated with changes in peel colour, acid content and pulp sugar (Lv *et al.*, 2023). Previous studies have reported the involvement of growth regulators such as auxin, cytokinin,

ethylene and abscisic acid (ABA) in citrus fruit maturation. These plant hormones regulate genes encoding carotenoids and chlorophyllase (chlorophyl breakdown) that are mainly responsible for the maturity-related phenotypes during fruit ripening (Lv *et al.*, 2023; McAtee *et al.*, 2013). The genes primarily involved in fruit ripening were also found to be coupled with starch and sucrose metabolism, oxidation–reduction, transmembrane transport and gluconeogenesis, plant hormone signal transduction and oxidative phosphorylation in citrus (Terol *et al.*, 2019). A deletion in Chr3 of *Citrus clementina* was shown to down-regulate a gene encoding SEP-like MADS box protein, which is a negative regulator of ethylene biosynthesis, inducing early fruit ripening (Terol *et al.*, 2019). The deleted regions in IrM2 also might have candidate genes that are negative regulators of the ripening process. A lot of small INDELs that were not identified by structural variation tools also might be associated with fruit ripening, seed reduction and improved peel colour of IrM2.

Examining the genomes of IrM2 and Murcott cultivars has provided a deeper insight into the effects of gamma irradiation and the chromosomal aberrations that it induces. Based on these findings, we suggest that more detailed analyses focusing on specific genes related to traits of interest can be pursued in future research and improvements of citrus can be achieved through more targeted editing approaches. The identified structural variations should be further verified by cytological analysis, PCR amplification with primers flanking the breakpoints and other targeted loci, gene expression and morphology-based assessments such as pollen and ovule viabilities in the mutant and progenitor genotypes to confirm that they are true genetic variants. The expression data obtained from RNA-seq experiments should be integrated with genomic data in the future to determine whether genes within or adjacent to structural variant regions exhibit altered expression patterns and thereby affect transcription.

Materials and methods

Plant materials, genome sequencing, genome assembly and annotation

Fresh, young, immature leaves of Murcott and IrM2 were obtained from the arboretum, Department of Agriculture and Fisheries (DAF), Bundaberg, Queensland, Australia. Total genomic DNA and RNA were extracted from fine pulverized leaf tissues and DNA; RNA sequencing of both Murcott and IrM2 sequencing was performed as detailed in Nakandala *et al.*, 2023. The Hi-C sequencing was performed at Arima Genomics, Australian National University, Australia. The Hi-C reads were subjected to trimming at 0.01 quality limit, adapter trimming and trimming of five over-represented 3 bp molecular barcode UMI sequences and 2 dark bases from the 5' end of the paired-end reads in CLC genomic workbench 24.0.1, before the genome assembly. Contig assemblies were generated using PacBio high fidelity (HiFi) reads and Hi-C reads using the Hi-C integrated mode of Hifiasm (v. 0.19.8), with parameters --n-weight 6, --n -perturb 15 000 and --f-perturb 0.3 to improve phasing. Scaffolding of the two haplotypes was independently performed using BWA aligner + Arima mapping and SALSA scaffolding pipeline as detailed in Nakandala *et al.* (2024). Scaffolding was also performed using the YaHS scaffolding tool to verify the results (Zhou *et al.*, 2023). The Hi-C contact maps were generated using Juice box 2.17.

The scaffolds of the Murcott hap1 were further aligned with the *C. sinensis* diploid reference genome (cv. Bingtang), retrieved

from the Zenodo repository (https://zenodo.org/record/8016647.), using D-genies (https://dgenies.toulouse.inra.fr/) (Cabanettes and Klopp, 2018). The scaffolds of Murcott hap2 were assigned to pseudochromosomes based on the Murcott hap1 assembly. The scaffolds of IRM2 haplotypes were assigned to pseudochromosomes based on the final Murcott haplotype chromosome-level assemblies. The joining of the scaffolds corresponding to the reference assemblies was manually performed by adding N's in between them. The orientation of the pseudochromosomes was determined based on our previously published *C. australis* genome (Nakandala *et al*., 2023), to keep the genomes of all the Australian limes consistent. The structural and functional annotations of the genomes were performed as described in Nakandala *et al*. (2024) and Nakandala *et al*. (2023).

Identification of the translocated segment and the mapping of the IrM2 chromosomes and translocated segments to MT chromosomes

The SC9 (IrM2-hap1 chr6) was aligned with the Chr6 of MT-hap1 to visualize and identify the translocated region of Chr3. The nucleotide sequence of SC9, which was matching with the Chr6 of MT-hap1, was identified by the long-read support mapping tool with default parameters (Match score-2, mismatch cost-4, gap open cost-4, gap extend cost-2, long gap open cost-24, long gap extend cost-1) in CLC Genomics Workbench. The translocated segment of Chr3 (1st segment of SC9–3.5 Mb) was identified and extracted from the complete sequence of SC9 in Clone Manager Ver. 9. The alignments of the translocated segment and the Chr3 of IrM2 with chr3 of MT-hap1 were done using the above same mapping tool and parameters. The dot plots for these alignments were generated by D-genies using minimap2 with the same default parameters.

Structural variations identification between the two genomes

The structural similarities and discrepancies between the homologous chromosomes of the two genomes were predicted using Synteny and Rearrangement Identifier (SyRI) (Goel *et al*., 2019). Structural variants identified by SyRI were annotated with overlapping genes from the IrM2 genome using a custom Python script that was implemented using Python 3, with standard libraries including csv and collections. Here, a coordinate-based comparison was performed using the gene annotations from the gff3 file to identify genes whose genomic intervals overlap with variant regions. The code used to annotate SyRI variants with gene overlaps is available from the authors upon request.

The presence-absence variations between two whole genome assemblies (reference and query) were identified using the scanPAV tool. For that, first, the soft-masked reference assembly was shredded into 1 kb fragments after the removal of N bases. The fragments were then aligned with the soft-masked query assembly using the Burrows-Wheeler Aligner (BWA) tool. The small repeats were filtered, and the mapping coordinates were identified. The 1 kb fragments that were only present in the reference and absent in the query assembly were extracted. The closely located sequences were merged into a single sequence. The obtained present sequences in the reference assembly were further filtered for sequences greater than 1000 bp to avoid noise using the tool; filter sequences by length in Galaxy version 1.2 (https://usegalaxy.org.au/). The filtered sequences were then dot plotted against the same reference genome, and the query genome independently using D-genies to further confirm that those sequences were only present in the reference; however, they were absent in the query. The sequences that were only present in the query genome were also identified in the same way.

Accession numbers

The whole genome sequence data for Murcott haplotype assemblies have been deposited to NCBI GenBank sequence database under accession numbers JBFPLR000000000, JBFPLS000000000, bioprojects [PRJNA1127039, PRJNA1127040] and biosample [SAMN37218425]. The whole genome sequence data for IrM2 haplotype assemblies have been deposited to NCBI GenBank sequence database under accession numbers JBMGNQ000000000, JBMGNP000000000, bioprojects [PRJNA 1127058, PRJNA1127059] and biosample [SAMN41763727]. The whole genome sequence data and annotation data for Murcott and IrM2 haplotype assemblies have been deposited in the Genome Warehouse in National Genomics Data Center (Chen *et al*., 2021; CNCB-NGDC 2022), Beijing Institute of Genomics, Chinese Academy of Sciences/China National Center for Bioinformation, under accession numbers GWHEUEF00000000.1, GWHEUEG00000000.1, GWHEUEJ00000000.1 and GWHEUEK00000000.1, BioProjects [PRJCA027797 and PRJCA027802] and Biosamples [SAMC3990319 and SAMC3990366] that are publicly accessible at https://ngdc.cncb.ac.cn/gwh. The whole genome and annotation data have also been submitted to Citrus genome database, under the accession numbers CGD24004.1 (Murcott-hap1), CGD24004.2 (Murcott-hap2), CGD24005.1 (IrM2-hap1) and CGD24005.2 (IrM2-hap2) (https://www.citrusgenomedb.org/).

Author contributions

The authors confirm contribution to the paper as follows: Study conception, design and supervision: Henry R, Furtado A, Kharabian Masouleh A; sample collection: Nakandala U, Mason P, Furtado A, Smith M; advice on laboratory experiments and sample collection: Furtado A, Henry R; data analysis and interpretation of results: Nakandala U, Henry R, Furtado A, Kharabian Masouleh A; draft manuscript preparation including all the figures, tables and supplementary materials: Nakandala U; all authors reviewed the results and approved the final version of the manuscript.

Queensland and contributions from the Australian Government and Bioplatforms Australia.

Conflict of interest

The authors declare no competing interests.

Data availability statement

The data that support the findings of this study are openly available in NCBI at https://ncbi.nlm.nih.gov, reference number JBFPLR000000000 and JBFPLS000000000.

References

Barry, G.H., Caruso, M. and Gmitter, F.G., Jr. (2020) Commercial scion varieties. In *The Genus Citrus* (Talon, M., Caruso, M., Fred, G. and Gmitter, J., eds), pp. 83–104. Amsterdam, Netherlands: Elsevier.

Bermejo, A., Pardo, J. and Cano, A. (2011) Influence of gamma irradiation on seedless citrus production: pollen germination and fruit quality. *Food Nutr. Sci.* **2**, 169–180.

Bosco, G. and Haber, J.E. (1998) Chromosome break-induced DNA replication leads to nonreciprocal translocations and telomere capture. *Genetics* **150**, 1037–1047.

Cabanettes, F. and Klopp, C. (2018) D-GENIES: dot plot large genomes in an interactive, efficient and simple way. *PeerJ* **6**, e4958.

Caruso, M., Las Casas, G., Scaglione, D., Gattolin, S., Rossini, L., Distefano, G., Cattonaro, F. et al. (2019) Detection of natural and induced mutations from next generation sequencing data in sweet orange bud sports. *Acta Hortic* **1230**, 119–124. https://doi.org/10.17660/actahortic.2019.1230.15

Caruso, M., Smith, M.W., Froelicher, Y., Russo, G. and GmitterJr, F.G. (2020) Traditional breeding. In *The Genus Citrus* (Talon, M., Caruso, M. and Gmitter, F.G., eds), pp. 129–148. Cambridge: Woodhead Publishing.

Chen, M., Ma, Y., Wu, S., Zheng, X., Kang, H., Sang, J., Xu, X. et al. (2021) Genome warehouse: A public repository housing genome-scale data. *Genomics, Proteomics & Bioinformatics* **19**, 584–589. https://doi.org/10.1016/j.gpb.2021.04.001

Farré, A., Cuadrado, A., Lacasa-Benito, I., Cistué, L., Schubert, I., Comadran, J., Jansen, J. et al. (2012) Genetic characterization of a reciprocal translocation present in a widely grown barley variety. *Mol. Breed.* **30**, 1109–1119.

Forejt, J. (2001) Nondisjunction. In *Encyclopedia of Genetics* (Brenner, S. and Miller, J.H., eds), pp. 1345–1347. Cambridge, MA: Academic Press.

Goel, M., Sun, H., Jiao, W.-B. and Schneeberger, K. (2019) SyRI: finding genomic rearrangements and local sequence differences from whole-genome assemblies. *Genome Biol.* **20**, 1–13.

Hardy, S., Barkley, P., Treeby, M., Smith, M. and Sanderson, G. (2017) *Australian Mandarin Production Manual*. Parramatta, NSW: NSW Department of Primary Industries.

Hase, Y., Satoh, K. and Kitamura, S. (2023) Comparative analysis of seed and seedling irradiation with gamma rays and carbon ions for mutation induction in Arabidopsis. *Front. Plant Sci.* **14**, 1149083.

Haskins, J.S. and Kato, T.A. (2019) Reciprocal Translocation Analysis with Whole Chromosome Painting for FISH. *Methods Mol Biol.* **1984**, 117–122.

Komura, S., Jinno, H., Sonoda, T., Oono, Y., Handa, H., Takumi, S., Yoshida, K. et al. (2022) Genome sequencing-based coverage analyses facilitate high-resolution detection of deletions linked to phenotypes of gamma-irradiated wheat mutants. *BMC Genomics* **23**, 1–21.

Leteurtre, F., Li, X., Gluckman, E. and Carosella, E. (1997) Telomerase activity during the cell cycle and in gamma-irradiated hematopoietic cells. *Leukemia* **11**, 1681–1689.

Li, F., Shimizu, A., Nishio, T., Tsutsumi, N. and Kato, H. (2019) Comparison and characterization of mutations induced by gamma-ray and carbon-ion irradiation in rice (Oryza sativa L.) using whole-genome resequencing. *G3 (Bethesda)* **9**, 3743–3751.

Lv, Y., Ren, S., Wu, B., Jiang, C., Jiang, B., Zhou, B., Zhong, G. et al. (2023) Transcriptomic and physiological comparison of Shatangju (Citrus reticulata)

and its late-maturing mutant provides insights into auxin regulation of citrus fruit maturation. *Tree Physiol.* **43**, 1841–1854.

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

Matsuda, Y., Uchimura, A., Satoh, Y., Kato, N., Toshishige, M., Kajimura, J., Hamasaki, K. et al. (2023) Spectra and characteristics of somatic mutations induced by ionizing radiation in hematopoietic stem cells. *Proc. Natl. Acad. Sci.* **120**, e2216550120.

Mba, C. and Shu, Q. (2012) Gamma irradiation. In *Plant Mutation Breeding and Biotechnology* (Shu, Q.Y., Forster, B.P. and Nakagawa, H., eds), pp. 91–98. Wallingford UK: CABI.

McAtee, P., Karim, S., Schaffer, R. and David, K. (2013) A dynamic interplay between phytohormones is required for fruit development, maturation, and ripening. *Front. Plant Sci.* **4**, 79.

Mennuti, M. (2019) Cytogenetics: Part 2, Structural Chromosome Rearrangements and Reproductive Impact. In *Perinatal Genetics* (Norton, M., Kuller, J. and Dugoff, L., eds), pp. 39–43. Missouri: Elsevier.

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

Nakandala, U., Masouleh, A.K., Smith, M.W., Furtado, A., Mason, P., Constantin, L. and Henry, R.J. (2023) Haplotype resolved chromosome level genome assembly of Citrus australis reveals disease resistance and other citrus specific genes. *Horticult. Res.* **10**, uhad058.

Nakandala, U., Furtado, A., Masouleh, A.K., Smith, M.W., Williams, D.C. and Henry, R.J. (2024) The genome of Citrus australasica reveals disease resistance and other species specific genes. *BMC Plant Biol.* **24**, 260.

Riviello-Flores, M.d.l.L., Cadena-Iñiguez, J., Ruiz-Posadas, L.d.M., Arévalo-Galarza, M.d.L., Castillo-Juárez, I., Soto Hernández, M. and Castillo-Martínez, C.R. (2022) Use of gamma radiation for the genetic improvement of underutilized plant varieties. *Plan. Theory* **11**, 1161.

Smith, M.W. (2006) Variety 'IrM2'. *Plant Var. J.* **19**, 260–264 Available from: www.ipaustralia.gov.au/tools-and-research/professional-resources/ip-rights-journals/plant-varieties-journals

Terol, J., Nueda, M.J., Ventimilla, D., Tadeo, F. and Talon, M. (2019) Transcriptomic analysis of Citrus clementina mandarin fruits maturation reveals a MADS-box transcription factor that might be involved in the regulation of earliness. *BMC Plant Biol.* **19**, 1–20.

Tian, S., Ge, J., Ai, G., Jiang, J., Liu, Q., Chen, X., Liu, M. et al. (2021) A 2.09 Mb fragment translocation on chromosome 6 causes abnormalities during meiosis and leads to less seed watermelon. *Hortic. Res.* **8**, 1–13.

Verma, R.C., Purbiya, R. and Khah, M.A. (2019) Gamma Irradiation Induced Reciprocal Translocation in Pea (Pisum sativum L.). *Chromosome Bot.* **13**, 71–74.

Zhou, C., McCarthy, S.A. and Durbin, R. (2023) YaHS: yet another Hi-C scaffolding tool. *Bioinformatics* **39**, btac808.

Zhu, Y. (2018) *Genomic Differences Between W. Murcott Mandarin and Its Mutational Derivative Tango*. UC Riverside.

Supporting information

Additional supporting information may be found online in the Supporting Information section at the end of the article.

Figure S1 The two mandarin varieties: Murcott and IrM2.

Figure S2 The alignment of the scaffolds of MT-hap1 assembly against the *C. sinensis* haplotypes and the Hi-C contact map for MT-hap1 assembly.

Figure S3 The assignment of the MT-hap2 scaffolds into pseudochromosomes.

Figure S4 The alignment of IrM2-hap1 scaffolds against Murcott haplotype assemblies and the Hi-C contact map for IrM2-hap1 assembly.

Figure S5 The alignment of IrM2-hap2 scaffolds against Murcott haplotype assemblies and the Hi-C contact map for IrM2-hap2 assembly.

Figure S6 The structural similarities and differences between the two Murcott haplotype assemblies.

Figure S7 The structural similarities and differences between MT-hap2 and IrM2-hap1 assemblies.

Figure S8 The structural similarities and differences between MT-hap1 and IrM2-hap2 assemblies.

Figure S9 Structural rearrangements between IrM2 haplotypes.

Figure S10 The alignment of the translocated fragment of Chr3 (SC9 1st part) and the rest of the segment of Chr3 (SC3 and SC12) independently with the Chr3 of Mt-hap1 assembly.

Figure S11 The sequences in Chr3 that were only present in MT-hap1 and absent in IrM2-hap1, identified by the scanPAV tool.

Figure S12 The mapping of the SC9, comprised of the translocated segment (1st segment), and the rest of the SC9 (2nd segment) with the Chr6 of MT-hap1 chromosome.

Table S1 MT-hap1 scaffolds corresponding to *C. sinensis* chromosomes.

Table S2 MT-hap2 scaffolds corresponding to MT-hap1 chromosomes.

Table S3 IrM2-hap1 scaffolds corresponding to MT-hap1 and hap2 chromosomes.

Table S4 IrM2-hap2 scaffolds corresponding to MT-hap1 and hap2 chromosomes.

Table S5 The functional annotation details of Murcott haplotypes.

Table S6 The functional annotation details of IrM2 haplotypes.

Table S7 The genes that were only present in the Chr3 of MT-hap1 1 (0.163 Mb) and were absent in IrM2-hap1, that were identified by the scanPAV analysis.

Table S8 The genes that were only present in the Chr7 of MT-hap2 and were absent in IrM2-hap2, that were identified by the scanPAV analysis.

Table S9 Number of duplicated, inverted duplicated regions and copy losses in individual chromosomes of IrM2.

Table S10 The genes in IrM2-hap1 that are associated with inverted duplicates, translocations and inverted translocations.

Table S11 The genes in IrM2-hap2 that are associated with inverted duplicates, translocations and inverted translocations.