Identifying radiation induced mutations in commercial low seed Murcott mandarins.

Agri-Science Queensland Innovation Opportunity

2018

Matthew Webb

and

Natalie Dillon

Subtropical Fruits and Genetic Improvement

Horticulture and Forestry Science

This publication has been compiled by Matthew Webb and Natalie Dillon, Horticulture and Forestry Science, Department of Agriculture Fisheries and Forestry.

Acknowledgements: The authors would like to thank Malcolm Smith, Horticulture and Forestry Science, Department of Agriculture Fisheries and Forestry for provision of plant material.

Summary

Murcott and Afourer mandarin varieties are anticipated to drive future expansion of the mandarin export industry in Australia. The Department of Agriculture and Fisheries and commercial citrus growers have previously secured plant breeders rights for the propagation of low seed Murcott varieties generated via radiation mutagenesis. In this investigation, we sought to identify radiation induced mutations within three low seed Murcott varieties (IrM1, IrM2 and Phoenix) that could be used to effectively differentiate each cultivar. In the course of this investigation, we developed improved methods for isolating high quality citrus DNA up to 130 kb in size. We also gained expertise in the use of cutting-edge portable DNA sequencers which have broad utility in a range of diagnostic applications. Our investigations led to the development of useful software for processing long-read sequencing data and generated significant amounts of genomic information for the economically important Murcott variety. Future investigations will focus on optimising genome assembly algorithms and generating effective pipelines for identifying mutations within both long and short-read sequencing data. Ultimately, our aim is to develop molecular tests that can accurately identify low seed Murcott cultivars for variety protection purposes. Information gained from the investigation may also allow for the elucidation of genetic mechanisms underlying the low seed phenotype.

Table of contents

Table of figures

Table of tables

Background

Ionizing radiation is an important tool for crop improvement that has been used to develop more than 3200 new varieties since its first reported use in 1928 (FAO/IAEA 2014, Oladosu *et al.* 2015). Genetic variation achieved using this method ultimately arises from DNA damage and is associated with a variety of detectable mutations ranging from single base substitutions and small insertions/deletions (Du *et al.* 2017) to large deletions and chromosomal rearrangements (Behjati *et al.* 2016, Naito *et al.* 2005, Hase *et al.* 2018).

Despite massive improvements in genome sequencing technologies and data analysis techniques in recent years (Torkamaneh *et al.* 2018), accurate identification of mutations associated with important phenotypic traits remains challenging. In particular, detection of large structural variants and complex re-arrangements is difficult when using predominant short-read sequencing technologies (Tattini *et al.* 2015). More recently, however, long-read single-molecule sequencing technologies have emerged which provide new opportunities for their analysis (Sedlazeck *et al.* 2017). One particularly interesting long-read sequencing platform that has successfully been used for this purpose is the portable MinION DNA sequencer (Jain *et al.* 2017).

Mandarins were estimated to account for 23 million AUD worth of Australian exports between April and June 2017 and an increase in production is expected in coming years, especially for the Murcott and Afourer varieties (Hort Innovation, 2017). As with many fruit crops, seedlessness is a highly desirable trait in citrus cultivars and can be effectively induced using ionizing radiation (Sutarto *et al*., 2009). In Australia, low-seeded mutants have previously been derived from the Murcott variety via gamma irradiation and have been granted plant breeders rights for variety protection purposes (Queensland DPI 2003, Pressler 2004, Queensland DPI&F 2006). Despite receiving this protection, however, it can be difficult to differentiate various low seed mandarin varieties from each other.

In this investigation, we sought to use both short and long-read sequencing technologies to characterize radiation-induced mutations within three low-seeded Murcott mandarin varieties. It is anticipated that these investigations will allow for the development of genetic tests for variety protection purposes and potentially allow for characterisation of important regions involved in seed formation.

Project Objectives

The objectives of this investigation were to:

1. Optimise experimental methodologies for short-read sequencing of citrus DNA and long-read sequencing using the portable MinION sequencer platform.
2. Identify radiation-induced DNA mutations within three distinct low-seeded Murcott varieties that could allow for their unique identification.

Ultimately, it is anticipated that this information will be used for the development of rapid, cost-effective molecular tests for variety protection purposes.

Methodology

Plant material

Two low-seed mandarin varieties, ‘IrM1’ and ‘IrM2’, that were developed within the Queensland Department of Agriculture and Fisheries via gamma irradiation of ‘Murcott’ bud sticks have been previously described (Queensland DPI 2003, Queensland DPI&F 2006). Another radiation-induced, low-seed variety, ‘Phoenix’ (alternatively called ‘Code 66-75’), developed by a commercial grower (2PH Farms), has also been described (Pressler 2004).

DNA extraction

For short-read sequencing, DNA was extracted from leaf material essentially as described by Healy *et al* (2014). For long-read sequencing, a range of methods were attempted. These included modifications of the CTAB protocol used for short-read sequencing, a PowerSoil DNA Isolation kit (Mo Bio), and a Nucleon Phytopure kit (Illustra).

In order to remove small RNA fragments from RNase treated nucleic acid extracts, PEG/NaCl based DNA precipitation (He *et al.* 2013) or various concentrations of AmpureXP beads (Agencourt) were used. Several methods were also attempted to selectively remove polysaccharides from DNA extracts, including the use of various CTAB and salt concentrations as outlined by Darby *et al* (1970) and various salt and ethanol combinations as outlined by Fang *et al* (1992).

Short-read sequencing

Barcoded libraries were prepared from mandarin DNA and 2x150bp reads were subsequently generated on one lane with the Illumina HiSeq 4000 platform using the services of the Australian Genome Research Facility (AGRF). Short reads from the Illumina platform were inspected and trimmed essentially as described by de Vries *et al* (2018). Reads were aligned to available *Citrus sinensis* (Xu *et al.* 2012) and *Citrus clementina* (Wu *et al.* 2014) genome sequences using BWA (Li and Durbin, 2009).

Long-read sequencing

Long-read sequencing was performed using an Oxford Nanopore Technologies (ONT) MinION portable sequencer. Libraries were prepared using either an ONT Rapid Sequencing kit (SQK-RAD004) or 1D² Sequencing kit (SQK-LSK308). Libraries were sequenced on either an R9.5, R9.4 or R9.4.1 flow cell as required. Reads were base-called, trimmed, filtered and assembled essentially as described by Gautier *et al* (2018). A custom script based on a complexity algorithm by Wootton *et al* (1993) was developed for filtering out long artefact sequences obtained with the 1D² Sequencing kit.

Results

Isolation of high quality citrus DNA

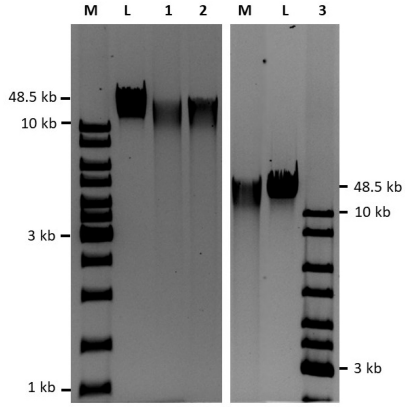
Isolation of significant amounts of high quality DNA was critical for our investigations. Towards this aim, we tested a range of commercial kits and published protocols on immature leaf samples. We found that a large scale CTAB protocol followed by RNase digestion was effective for isolating sufficient amounts of intact, total nucleic acid (TNA) with acceptable A260/A230 ratios (Figure 1). However, subsequent attempts to remove contaminating RNA fragments from these preparations via PEG precipitation resulted in significantly reduced A260/230 ratios. Whilst these extracts were found to be suitable for short-read sequencing on the Illumina platform, it was found that removal of RNA fragments using carboxylated magnetic beads resulted in higher purity and was necessary for effective long-read sequencing on the MinION platform.

In order to potentially detect mutations induced by ionizing radiation within each low-seed mandarin variety, we performed short-read sequencing on DNA extracts using the Illumina platform. Following deduplication and quality trimming, we found that we were able to achieve average coverage levels of approximately 100x when mapped to either *Citrus clementina* or *C. sinensis* chromosomal scaffolds (Table 1). For both reference genomes, major scaffolds representing chromosomes 1-9 were

observed to be significantly lower in total size compared to their estimated total length of ~370 Mb, presumably due to incomplete sequencing. Our investigations revealed that reads from each of our samples covered approximately 96% and 90% of the *C. clementina* and *C. sinensis* genomes respectively.

Figure 1 – Purity and integrity of nucleic acids isolated from citrus.

A



Lane	Description
M	250 ng Generuler 1 kb DNA ladder
L	250 ng Lambda DNA (48.5 kb).
1	83 ng IrM2 TNA+RNase. CTAB extraction method.
2	96 ng Murcott DNA. CTAB→ beads extraction method.
3	87 ng IrM1 DNA. CTAB→ PEG extraction method.

B

Sample	Extract Type	Extraction Method	A260/280	A260/230
IrM2	TNA+RNase	CTAB	2.01	2.32
IrM1	DNA	CTAB→ PEG	1.78	1.36
Murcott	DNA	CTAB → beads	1.79	2.21

A - 0.8% agarose gel electrophoresis of nucleic acids.

B - Spectrophotometric analysis of nucleic acid purity.

Table 1 – Coverage of short-read data against citrus reference genomes

Reference Genome	Total length of chromosome 1-9 scaffolds (GB)	Sample Name	Average Coverage	Percent bases covered[#]
Citrus clementina	288.6	Murcott	105.71	95.61
		IrM1	113.90	95.93
		IrM2	105.71	95.91
		Phoenix	111.76	95.94
Citrus sinensis	239.0	Murcott	98.58	90.00
		IrM1	105.12	89.99
		IrM2	97.18	89.96
		Phoenix	102.64	90.01

[#] Coverage levels greater than or equal to 1

Long-read sequencing

In order to potentially improve the detection of large structural variants, we sought to obtain long-read sequences using a portable MinION sequencer recently developed by Oxford Nanopore Technologies (ONT). This approach proved to be technically challenging and revealed stringent nucleic acid purity and integrity requirements. Overall, we obtained the highest raw sequencing yields using an SQK-LSK308 1D² Sequencing Kit (Table 2). With this kit, we were able to obtain 464,115 passed reads from a wild type Murcott DNA extract and achieved satisfactory read length metrics (N50=6431 bp).

However, data obtained with the 1D² sequencing kit was also associated with very long sequencing artefacts that had to be removed using an additional low-complexity filter.

The longest read observed in our study was approximately 130 kb and was obtained with an SQK-RAD004 Rapid 1D sequencing kit (Table 2) and wild-type Murcott DNA. Subsequent analysis of the longest 1D read by BLAST revealed that it shared multiple regions with 90% identity to a section of chromosome 1 from *Citrus clementina*. This level of homology appears to accord with an expected 94% accuracy rate reported for 1D reads (Tyler *et al.* 2018) and the previously reported genetic similarity of these cultivars (Wu *et al.* 2014). However, it is significant these regions of homology only covered 68% of the corresponding region from *C. clementina* and were interspersed with non-homologous regions.

Yields obtained with Rapid 1D sequencing kits were generally about half that observed with the ligation-based 1D² Sequencing Kit (SQK-LSK308). In the case of 1D² sequencing, both DNA strands are preferentially sequenced in order to obtain up to 97% accuracy. However, due to time limitations and complexities associated with basecalling and processing 1D² reads, we processed them as 1D reads during subsequent assembly stages.

Table 2 – Long-read sequencing of citrus DNA using the portable MinION sequencer.

Flow Cell	Sample Name	Library Kit	DNA Extraction Method	Input DNA (µg) [&]	Available Pores	Run Time (hr)	Sequences Passed				
							%	Number	N50[#] (bp)	Longest (bp)	Total Yield (GB)
R9.5	Murcott	SQK-RAD004	CTAB→PEG	0.61	1291	2.50	41	6002	5105	41620	0.18
	Murcott *	SQK-LSK308	CTAB→beads	8.00	782	50.00	45	403052 [402356]	5723 [5707]	678359 [85837]	1.28 [1.28]
R9.5	Murcott	SQK-LSK308	CTAB→beads	7.73	1290	45.75	36	464339 [464115]	6437 [6431]	603775 [96212]	1.94 [1.94]
R9.4.1	Murcott	SQK-RAD004	CTAB→beads	6.10	1486	15.50	44	237780	5907	130274	0.73
	Murcott *	SQK-RAD004	CTAB→beads	9.00	297	32.50	83	67038	6578	96335	0.22
R9.4	IrM1	SQK-RAD004	CTAB→PEG	1.82	918	3.25	87	13637	4880	62992	0.37
	IrM1 *	SQK-RAD004	CTAB→beads	6.21	677	54.50	63	188120	1663	49705	0.24
R9.4	IrM2	SQK-RAD004	CTAB→beads	5.99	1490	2.75	38	157651	9925	111484	0.69
	IrM2 [$]	SQK-RAD004	CTAB→beads	0.59	624	43.25					

* Library loaded on re-used, washed flow cells.
$ Library loaded on an already running flow cell.
& Input DNA amount for library construction.
# Sequences greater than this length constitute 50% of all passed bases.
[ ] Values calculated after filtering out low-complexity artefacts.

Overall, we obtained relatively low total sequencing yield for samples IrM1 and IrM2 (0.59 GB and 0.69 GB of passed reads). In addition, a low N50 value for one of the IrM1 DNA extracts (1663 bp) indicates that DNA degradation was a particular problem for this sample. By using multiple flow cells for the wild-type Murcott sample, we were able to obtain 4.45 GB of passed sequence data to assist in subsequent assembly of a reference genome.

Long read assembly

In order to assemble our long-read data into contigs, we used the correction and trimming modules of Canu (Koren *et al.* 2017), followed by assembly with SMARTdenovo. The longest contig assembled using this pipeline was a 461 kb region of chromosome 1 from wild-type Murcott mandarin (Table 3). It is interesting to note that this contig did not contain the longest 130 kb read detailed in Table 2 but possessed a similar level of coverage (63%) against chromosome 1 of *C. clementina*. Assembly

metrics (N50 values and total bases assembled) indicate that approximately 20 percent of the complete wild-type Murcott genome (~370 Mb) was assembled into contigs greater than 45 kb in length.

For both IrM1 and IrM2, the largest assembled contig was ~300 kb in size (Table 3) and aligned to the recently published 641 kb mitochondrial genome of *Citrus sinensis* (Yu *et al.* 2018). Further analysis (data not shown) indicated that other overlapping contigs derived from the mitochondrial genome were in fact present in the data, thus indicating that improvements in the assembly algorithm can be attained.

Table 3 – Assembly of long-read data from the MinION platform.

Sample Name	No. Contigs	No. Contigs >50 kb	Longest Contig (bp)	N50[#] (bp)	Total Bases (MB)*	Best BLAST Hit of Longest Contig		
						Description	Coverage (%)	Identity (%)
Murcott	4834	832	460791	45533	155.85	*C. Clementina* chromosome 1	63	96
IrM1	23	2	312471	199243	0.72	*C. sinensis* mitochondrial genome	93	96
IrM2	102	3	322706	24703	2.29	*C. sinensis* mitochondrial genome	93	98

[#] Contigs greater than this length constitute 50% of all assembled bases.
***** Bases contained in unassembled singletons are not included in this value.

Conclusions/Significance/Recommendations

In this investigation, we sought to characterise radiation induced mutations within three low-seeded Murcott mandarin varieties (IrM1, IrM2 and Phoenix). Ultimately, our investigation aims to facilitate the development of diagnostic tests that can differentiate each cultivar for variety protection purposes. Our strategy for mutation detection involved the use of both short-read sequencing techniques, which are relatively well established, and more recently developed long-read sequencing techniques (Stancu *et al.* 2017).

During our investigation, we found that pure, high-molecular weight DNA which does not possess any traces of degradation is critical for effective long-read sequencing using MinION portable DNA sequencers. Other researchers have previously developed methods to obtain high quality DNA from citrus (Terol *et al.* 2015, Shimizu *et al.* 2016), however they are laborious and have not been specifically tested on the MinION platform. The method developed in this investigation was comparatively easy and successfully isolated DNA strands up to 130 kb in size.

In recent reports, reads as long as 2.3 Mb (Payne *et al.* 2018) and raw sequencing outputs as high as 30 GB per flow cell have been achieved with MinION devices. However, we typically achieved results that were an order of magnitude lower than this. It therefore appears that significant opportunities remain to improve our methodologies. Our investigations did reveal that significantly higher yields can be obtained with ligation-based library preparation kits compared to transposase-based kits although the latter are considerably more convenient and have been used to provide longer read lengths. Use of mechanical shearing to increase fragment uniformity, or gel purification steps to remove smaller interfering DNA fragments represent feasible options for us to potentially improve our results (Schalamun *et al.* 2018).

Citrus clementina is the closest relative of Murcott mandarin that currently possesses a published, assembled genome (Wu *et al.* 2014). Initial analysis of our aggregated short-read data suggest that the Murcott genome is more than 95% identical to *C.clementina*. This level of homology should allow for accurate identification of small insertions, deletions and substitutions in mutant genomes using *C. clementina* as a reference. We also appear to have sufficient short-read data to potentially allow for identification of larger deletions based on coverage analysis. In subsequent investigations, our broad strategy is to identify mutations relative to the *C. clementina* reference genome that are uniquely observed within each low seed mutant and are absent within wild-type Murcott DNA.

For our unassembled long-read data, preliminary alignments revealed extended regions of approximately 90 percent identity between Murcott DNA and the *C. clementina* reference genome. Moreover, initial assembly attempts using only long-read data resulted in improved identity levels of 96 percent. However, in both cases, the regions of homology appeared to be interrupted by non-homologous stretches and overall coverage levels were less than 70 percent. These results suggest the existence of extended, low-quality regions within the long reads. Further analysis is required to confirm this and determine if alternative base-calling algorithms can be used to improve long-read quality. In particular, we will seek to determine if base-calling algorithms available for $1D^2$ sequencing reads provide a useful advantage for our purposes when compared to 1D reads.

Algorithm design for efficient assembly of error-prone, long-read sequence data remains an active area of development (Koren *et al.* 2017, Li 2016). In this study, we used a pipeline that assembled approximately 20 percent of the total estimated genome (~370 Mb) of wild-type Murcott into contigs greater than 45kb in length and achieved a maximum contig length of 460 kb. However, our analyses indicated that complete assembly was not always achieved and that further optimisation, or use of alternative algorithms, could significantly enhance the assembly process. We currently possess a significant amount of unassembled long and short-read data for wild-type Murcott and anticipate that hybrid assemblies will allow for the generation of useful reference sequences from this cultivar. It remains to be determined if we have sufficient long-read data from IrM1 and IrM2 to assist in the identification of large structural variants and chromosomal re-arrangements.

The low seed mutants examined in this study were vegetatively propagated from multi-cellular bud tissue exposed to gamma radiation. Given this background, it is possible that these cultivars will possess chimeric tissues harbouring various mutant genotypes (Frank and Chitwood, 2016). Consequently, we plan to use algorithms for mutant identification that can tolerate deviations from fixed ploidy levels. Importantly, previous studies have used similar strategies to ours to identify mutations within irradiated citrus tissues using short-read sequencing data (Terol *et al.* 2015) and a research group recently used this approach to successfully develop molecular tests for citrus cultivar identification (Las Casas *et al.* 2018). We expect similar results using short read data from this investigation and anticipate gaining further insights following integration of our long-read sequencing data.

Key Messages

Murcott and Afourer varieties are anticipated to drive future expansion of the mandarin export industry in Australia. The current investigation provides an important basis for the development of diagnostic tests that can conclusively identify valuable mandarin varieties developed by both DAF and commercial growers for variety protection purposes. As a result of this investigation, we were able to develop improved methods for isolating high molecular weight DNA from citrus plants and have

subsequently shown that these methods can be applied to other citrus varieties and crop species. We additionally gained valuable experience in the use of cutting-edge portable DNA sequencers which have broad utility in a range of novel applications. We also, developed useful software for processing long-read sequencing data and established pipelines for large-scale data analysis. It is anticipated that technologies developed in this project can be applied not only to various citrus varieties but numerous other agronomically important crop species.

Where to next

Future research efforts will focus on computer-based analyses of our data so that radiation induced mutations unique to each mandarin variety can be characterised. Good progress has already been made towards this goal using available short-read sequence data and established algorithms. More effort will be required to optimise techniques for mutation detection in the long-read sequencing data. In particular, appropriate software will need to be selected for performing hybrid assemblies and identifying somatic mutations from long-reads. Sufficient data has been generated in this investigation for generating publications and will form a valuable resource for future studies of related citrus varieties. Ultimately, further funding will be sought to optimise and validate molecular tests designed from our sequencing data for rapid identification of low seed Murcott varieties.

Budget Summary

A total of $8151.76 AUD was spent prior to the end of the 2017/2018 financial year. $6828.71 AUD was used to purchase flow cells, kits, devices, reagents and services for DNA sequencing. $1323.05 AUD was used for DNA extraction kits, reagents and consumables.

References

Behjati, S, Gundem, G, Wedge, DC, Roberts, ND, Tarpey, PS, Cooke, SL, Van Loo, P, Alexandrov, LB, Ramakrishna, M, Davies, H, Nik-Zainal, S, Hardy, C, Latimer, C, Raine, KM, Stebbings, L, Menzies, A, Jones, D, Shepherd, R, Butler, AP, Teague, JW, Jorgensen, M, Khatri, B, Pillay, N, Shlien, A, Futreal, PA, Badie, C, McDermott, U, Bova, GS, Richardson, AL, Flanagan, AM, Stratton, MR, Campbell, PJ & Group, IP 2016, 'Mutational signatures of ionizing radiation in second malignancies', *Nat Commun*, vol. 7, p. 12605.

Darby, GK, Jones, AS, Kennedy, JF & Walker, RT 1970, 'Isolation and analysis of the nucleic acids and polysaccharides from Clostridium welchii', *J Bacteriol*, vol. 103, no. 1, pp. 159-65.

de Vries, J, Curtis, BA, Gould, SB & Archibald, JM 2018, 'Embryophyte stress signaling evolved in the algal progenitors of land plants', *Proc Natl Acad Sci U S A*, vol. 115, no. 15, pp. E3471-E80.

Du, Y, Luo, S, Li, X, Yang, J, Cui, T, Li, W, Yu, L, Feng, H, Chen, Y, Mu, J, Chen, X, Shu, Q, Guo, T, Luo, W & Zhou, L 2017, 'Identification of substitutions and small insertion-deletions Induced by carbon-ion beam irradiation in *Arabidopsis thaliana*', *Front Plant Sci*, vol. 8, p. 1851.

Fang, G, Hammar, S & Grumet, R 1992, 'A quick and inexpensive method for removing polysaccharides from plant genomic DNA', *Biotechniques*, vol. 13, no. 1, pp. 52-4.

FAO/IAEA 2014, 'FAO/IAEA mutant variety database', Available at: https://mvd.iaea.org/ [Retrieved 31 July 2014].

Frank, MH & Chitwood, DH 2016, 'Plant chimeras: The good, the bad, and the 'Bizzaria'', *Developmental Biology*, vol. 419, no. 1, pp. 41-53.

Gautier, M, Yamaguchi, J, Foucaud, J, Loiseau, A, Ausset, A, Facon, B, Gschloessl, B, Lagnel, J, Loire, E, Parrinello, H, Severac, D, Lopez-Roques, C, Donnadieu, C, Manno, M, Berges, H, Gharbi, K, Lawson-Handley, L, Zang, LS, Vogel, H, Estoup, A & Prud'homme, B 2018, 'The Genomic Basis of Color Pattern Polymorphism in the Harlequin Ladybird', *Curr Biol*, doi: 10.1016/j.cub.2018.08.023

Hase, Y, Satoh, K, Kitamura, S & Oono, Y 2018, 'Physiological status of plant tissue affects the frequency and types of mutations induced by carbon-ion irradiation in Arabidopsis', *Sci Rep*, vol. 8, no. 1, p. 1394.

He, Z, Zhu, Y & Gu, H 2013, 'A new method for the determination of critical polyethylene glycol concentration for selective precipitation of DNA fragments', *Appl Microbiol Biotechnol*, vol. 97, no. 20, pp. 9175-83.

Healey, A, Furtado, A, Cooper, T & Henry, RJ 2014, 'Protocol: a simple method for extracting next-generation sequencing quality genomic DNA from recalcitrant plant species', *Plant Methods*, vol. 10, p. 21.

Hort Innovation 2017, 'Horticulture trade intelligence: Citrus: Q2 2017', Available at: https://horticulture.com.au/wp-content/uploads/2017/10/Q2_2017_Citrus_Trade-report.pdf [Accessed 15 October 2018].

Jain, M, Koren, S, Miga, KH, Quick, J, Rand, AC, Sasani, TA, Tyson, JR, Beggs, AD, Dilthey, AT, Fiddes, IT, Malla, S, Marriott, H, Nieto, T, O'Grady, J, Olsen, HE, Pedersen, BS, Rhie, A, Richardson, H, Quinlan, AR, Snutch, TP, Tee, L, Paten, B, Phillippy, AM, Simpson, JT, Loman, NJ & Loose, M 2018, 'Nanopore sequencing and assembly of a human genome with ultra-long reads', *Nat Biotechnol*, vol. 36, no. 4, pp. 338-45.

Jain, M, Olsen, HE, Paten, B & Akeson, M 2016, 'The Oxford Nanopore MinION: delivery of nanopore sequencing to the genomics community', *Genome Biol*, vol. 17, no. 1, p. 239.

Koren, S, Walenz, BP, Berlin, K, Miller, JR, Bergman, NH & Phillippy, AM 2017, 'Canu: scalable and accurate long-read assembly via adaptive *k*-mer weighting and repeat separation', *Genome Res*, vol. 27, no. 5, pp. 722-36.

Las Casas, G, Caruso, M, Scaglione, D, Gattolin, S, Rossini, L, Distefano, G, Romano, F, Cirrone, G, Cuttone, G, Cattonaro, F, Catara, A, Licciardello, G, Morganter, M & Licciardello, C 2018, 'Detection of natural and induced mutations from next generation sequencing data in sweet orange bud sports', *IV International Symposium on Citrus Biotechnology*, Uruguay, p. 83.

Li, H 2016, 'Minimap and miniasm: fast mapping and de novo assembly for noisy long sequences', *Bioinformatics*, vol. 32, no. 14, pp. 2103-10.

Li, H & Durbin, R 2009, 'Fast and accurate short read alignment with Burrows-Wheeler transform', *Bioinformatics*, vol. 25, no. 14, pp. 1754-60.

Naito, K, Kusaba, M, Shikazono, N, Takano, T, Tanaka, A, Tanisaka, T & Nishimura, M 2005, 'Transmissible and nontransmissible mutations induced by irradiating Arabidopsis thaliana pollen with gamma-rays and carbon ions', *Genetics*, vol. 169, no. 2, pp. 881-9.

Oladosu, Y, Rafii,MY, Abdullah, N, Hussin, G, Ramli, A, Rahim, HA, Miah, G & Usman, M 2016, 'Principle and application of plant mutagenesis in crop improvement: a review', Biotechnology and Biotechnological Equipment, vol. 30, no.1, pp.1-16.

Payne, A, Holmes, N, Rakyan, V & Loose, M 2018, 'Whale watching with BulkVis: A graphical viewer for Oxford Nanopore bulk fast5 files', *bioRxiv*, doi: 10.1101/312256.

Pressler, CR 2004, 'Code 66-75', *Plant Varieties Journal*, vol. 17, no. 3, pp. 155-158.

Queensland DPI 2003, 'IrM1', *Plant Varieties Journal*, vol. 16, no. 3, pp. 33-34.

Queensland DPI&F 2006, 'IrM2', *Plant Varieties Journal*, vol. 19, no. 3, pp. 260-264.

Schalamun, M, Nagar, R, Kainer, D, Beavan, E, Eccles, D, Rathjen, JP, Lanfear, R & Schwessinger, B 2018, 'Harnessing the MinION: An example of how to establish long - read sequencing in a laboratory using challenging plant tissue from *Eucalyptus pauciflora*', *Mol Ecol Resour*, doi: 10.1111/1755-0998.12938

Sedlazeck, FJ, Rescheneder, P, Smolka, M, Fang, H, Nattestad, M, von Haeseler, A & Schatz, MC 2018, 'Accurate detection of complex structural variations using single-molecule sequencing', *Nat Methods*, vol. 15, no. 6, pp. 461-8.

Shimizu, T, Kaminuma, E, Nonaka, K, Yoshioka, T, Goto, S, Matsumoto, T, Katayose, Y, Mochizuki, T, Tanizawa, Y, Toyoda, A, Fujiyama, A & Nakamura, Y 2016, 'A genomic approach to selecting robust and versatile SNP sets from next-generation sequencing data for genome-wide association study in citrus cultivars'*, Acta Hortic*, vol. 1135, pp. 23-32.

Stancu, MC, van Roosmalen, MJ, Renkens, I, Nieboer, MM, Middelkamp, S, de Ligt, J, Pregno, G, Giachino, D, Mandrile, G, Espejo Valle-Inclan, J, Korzelius, J, de Bruijn, E, Cuppen, E, Talkowski, ME, Marschall, T, de Ridder, J & Kloosterman, WP 2017, 'Mapping and phasing of structural variation in patient genomes using nanopore sequencing', *Nat Commun*, vol. 8, no. 1, p. 1326.

Sutarto, I, Agisimanto, D & Supriyanto, A 2009, 'Development of Promising Seedless Citrus Mutants through Gamma Irradiation', *Induced Plant Mutations in the Genomics Era. Food and Agriculture Organization of the United Nations, Rome*, pp. 306-308.

Tattini, L, D'Aurizio, R & Magi, A 2015, 'Detection of Genomic Structural Variants from Next-Generation Sequencing Data', *Front Bioeng Biotechnol*, vol. 3, p. 92.

Terol, J, Ibañez, V, Carbonell, J, Alonso, R, Estornell, LH, Licciardello, C, Gut, IG, Dopazo, J & Talon, M 2015, 'Involvement of a citrus meiotic recombination TTC-repeat motif in the formation of gross deletions generated by ionizing radiation and MULE activation', *BMC Genomics*, vol. 16, p. 69.

Torkamaneh, D, Boyle, B & Belzile, F 2018, 'Efficient genome-wide genotyping strategies and data integration in crop plants', *Theoretical and Applied Genetics*, vol. 131, no. 3, pp. 499-511.

Tyler, AD, Mataseje, L, Urfano, CJ, Schmidt, L, Antonation, KS, Mulvey, MR & Corbett, CR 'Evaluation of Oxford Nanopore's MinION Sequencing Device for Microbial Whole Genome Sequencing Applications', *Scientific Reports*, vol. 8, p. 10931.

Wootton, JC & Federhen, S 1993, 'Statistics of local complexity in amino acid sequences and sequence databases', *Computers & Chemistry*, vol. 17, no. 2, pp. 149-63.

Wu, GA, Prochnik, S, Jenkins, J, Salse, J, Hellsten, U, Murat, F, Perrier, X, Ruiz, M, Scalabrin, S, Terol, J, Takita, MA, Labadie, K, Poulain, J, Couloux, A, Jabbari, K, Cattonaro, F, Del Fabbro, C, Pinosio, S, Zuccolo, A, Chapman, J, Grimwood, J, Tadeo, FR, Estornell, LH, Muñoz-Sanz, JV, Ibanez, V, Herrero-Ortega, A, Aleza, P, Pérez-Pérez, J, Ramón, D, Brunel, D, Luro, F, Chen, C, Farmerie, WG, Desany, B, Kodira, C, Mohiuddin, M, Harkins, T, Fredrikson, K, Burns, P, Lomsadze, A, Borodovsky, M, Reforgiato, G, Freitas-Astúa, J, Quetier, F, Navarro, L, Roose, M, Wincker, P, Schmutz, J, Morgante, M, Machado, MA, Talon, M, Jaillon, O, Ollitrault, P, Gmitter, F & Rokhsar, D 2014, 'Sequencing of diverse mandarin, pummelo and orange genomes reveals complex history of admixture during citrus domestication', *Nature Biotechnology*, vol. 32, p. 656.

Xu, Q, Chen, LL, Ruan, X, Chen, D, Zhu, A, Chen, C, Bertrand, D, Jiao, WB, Hao, BH, Lyon, MP, Chen, J, Gao, S, Xing, F, Lan, H, Chang, JW, Ge, X, Lei, Y, Hu, Q, Miao, Y, Wang, L, Xiao, S, Biswas, MK, Zeng, W, Guo, F, Cao, H, Yang, X, Xu, XW, Cheng, YJ, Xu, J, Liu, JH, Luo, OJ, Tang, Z, Guo, WW, Kuang, H, Zhang, HY, Roose, ML, Nagarajan, N, Deng, XX & Ruan, Y 2013, 'The draft genome of sweet orange (Citrus sinensis)', *Nat Genet*, vol. 45, no. 1, pp. 59-66.

Yu, F, Bi, C, Wang, X, Qian X & Ye, N 2018, 'The complete mitochondrial genome of *Citrus sinensis', Mitochondrial DNA Part B*, vol. 3, no. 2, pp. 592-593.