Development of polymorphic simple sequence repeat (SSR) markers from genome re-sequencing of *Carica papaya* L. ‘Sunrise Solo’ and ‘RB2’ for marker-assisted breeding

U. Nantawan^{1,2}, C. Kanchana-udomkan^{1,2}, R. Drew¹ and R. Ford^{1,2}

¹ School of Natural Sciences, Griffith University, Nathan, Australia

²Environmental Futures Research Institute, Griffith University, Gold Coast, Australia

Abstract

Carica papaya L. is one of the top five tropical fruit crops grown worldwide and an important fruit crop in Australia. Expansion of the papaya industry in Australia is reliant on improved fruit qualities, together with consistent productivity. Although marker-assisted breeding has resulted in substantial genetic gains in other species, this approach in papaya is hampered not only by the subjective nature of many of the desirable target fruit quality traits but also by the dearth of polymorphic markers available, in part due to the narrow genetic base within the cultivated genome.

To identify a range of polymorphic sites for marker development and subsequent mapping, we performed whole-genome re-sequencing on two commercial varieties of papaya, ‘Sunrise Solo’ (Hawaiian variety) and ‘RB2’ (Australian variety), and developed highly polymorphic SSR markers. In total, 30.2 Gb and 32.4 Gb of genome sequence data were generated using Illumina HiSeq 4000 sequencing technology for ‘Sunrise Solo’ and ‘RB2’, respectively. These data represented approximately a 80-85× coverage of the papaya genome, which is estimated to be 372 Mb. Subsequently, 236 Mb of ‘Sunrise Solo’ and 239 Mb of ‘RB2’ unique sequences were anchored to the papaya reference genome (<https://genomevolution.org/>). The initially assembled genomes were compared and analysed for putative polymorphic microsatellite motifs. Microsatellite-containing coding sequences (CDS) were assigned to gene ontology (GO) terms for providing a set of polymorphic markers putatively associated with functional genes. Fifty of the selected microsatellites, with a broad range of putative function, revealed polymorphism as SSR markers on ‘Sunrise Solo’ and ‘RB2’ genotypes.

Keywords: Breeding, microsatellite, next-generation sequencing, *in silico* analysis, genome-wide polymorphism.

INTRODUCTION

Papaya (*Carica papaya L.*) is a fruit crop that has been cultivated commercially worldwide in tropical and sub-tropical regions. Annual production of papaya was approximately 64 million tons in 2014 (FAOSTAT, 2014). In Australia, the current annual yield of papaya is around 6.5 thousand tons (FAOSTAT, 2014) with strong potential for growth in future fruit markets. Improvement of fruit qualities together with consistent productivity has posed potential for an expansion of the papaya industry in Australia.

Among several commercial papayas, 'Sunrise Solo' and 'RB2' have superior fruit quality, and were chosen as key parental lines in fruit quality traits improvement program. Monitoring the genetic gains in their subsequent generations and selection based on directed accumulation of favoured alleles could make development of new commercial papaya varieties more efficient. It is likely to speed up the breeding process and reduce the number of plants which needs to be evaluated in the field trials. For this approach, the development of informative markers, which are easy to use and have high level of polymorphism between the two cultivars, is essential.

The use of molecular markers was reported as an effective tool for fruit quality trait selection in many fruit crops such as mango (Viruel et al., 2005; Dillon, 2013) and pear (Zhang et al., 2013). However, the availability of suitable markers has long been hampered by the laborious and costly process of marker development. In papaya, molecular markers have not yet been extensively used despite the availability of markers from genome map (Chen et al., 2007) and published papaya genome sequence (Ming et al., 2008). The development of practical markers associated with fruit quality traits remains limited by low level of polymorphism in breeding material, especially among cultivated genomes (Oliveira et al., 2010). As the result, there is a relatively small number of markers applicable for identification of target loci in elite material of papaya.

Among DNA marker types, simple sequence repeats (SSR) are a robust marker that have been widely used in papaya genotyping. This wide use is because SSRs have shown to be a reliable marker system and have analytical simplicity (Eustile et al., 2008 and Pinto et al., 2013). Several SSR studies were reported in papaya, including the development of 706 SSRs markers for a comprehensive genetic map (Chen et al., 2007). SSRs from papaya have been identified through the use of bacterial artificial chromosome (BAC) end sequences, expressed sequence tags (ESTs) and whole genome shotgun (WGS) sequences (Eustile et al., 2008; Wang et al., 2008 and Vidal et al., 2014). However, many of these SSRs developed

have a low polymorphic information content (PIC) value, approximately 17-23% (Wang et al., 2008 and Vidal et al., 2014), limiting their use in the investigated papaya genomes.

Recently, next-generation sequencing technology provides the possibility to identify genome-wide polymorphic loci, including microsatellites, supporting the development of locus-specific markers *in silico* (Wang et al., 2008 and Sun et al., 2013). The availability of the whole genome sequences of specific papaya genotypes will enable high-throughput detection of polymorphic regions. Moreover, associations between candidate SSR markers and functional genes can be determined via gene ontologies (Hennig et al., 2003; Walls et al., 2012), resulting highly informative SSR markers.

In this study, we demonstrate the use of sequencing technology and *in silico* polymorphism analysis for developing polymorphic SSR markers in papaya. A set of SSR markers obtained from genome-wide identification of SSR motifs within known functional genes can be used to efficiently discriminate between elite cultivars; 'Sunrise Solo' and 'RB2'. These SSR markers have potential for use in selective breeding strategies.

MATERIALS AND METHODS

Plant materials and genomic data

Genomic DNA was extracted by CTAB method (Delaporta et al., 1983) from the papaya genotypes 'Sunrise Solo' and 'RB2'. The two genomes were re-sequenced using the Illumina Hi-seq4000 platform. Subsequently, raw genome sequence data were filtered to obtain high quality reads, using FASTQ Groomer and Trimmomatic program on Galaxy Tool V 0.1 (<http://galaxy-ql.d.genome.edu.au>; University of Queensland Research Computing Centre, Australia). Then genome sequences were assembled by mapping to the papaya reference genome retrieved from <https://genomevolution.org>, using BWA software (Li and Durbin, 2009).

SSR detection and *in silico* polymorphism analysis

Microsatellite motifs were detected in the sequence assemblies using GMATo (Wang et al., 2013) and MISA scripts (<http://pgrc.ipk-gatersleben.de/misa>). The minimum length of motif was defined as 12 nucleotides and only SSR loci containing 1-6 bp repeat units were included. R software version 3.2.4 was used to identify location and length polymorphic SSR loci among the genotypes. Simultaneously, the loci were annotated and gene predictions performed based on papaya and *Arabidopsis thaliana* homology using Augustus, a plugin program within the Geneious 8 platform (Biomatters Ltd, New Zealand). In order to provide

a set of polymorphic markers associated with functional genes, marker-containing SSR sequences and identified coding sequences (CDS) were assigned Gene Ontology (GO) terms using B2GO software and the plant ontology tool (<http://bioinfo.cau.edu.cn/agriGO>).

Primer design and polymorphism analysis of DNA markers

For further validation, primer pairs were designed from SSR-flanking regions to produce amplicons sizes between 100 bp and 1200 bp, using Primer 3.0 (<http://primer3.ut.ee>). PCR reaction mixtures were performed in 20 µl and contained 1X My red Taq reaction buffer with Mg²⁺ and dNTP (Takara), 0.3 µM forward and reverse primers, 0.4 unit of Taq DNA polymerase and 50 ng of template DNA. DNA amplification was programmed with an initial denaturation at 94°C for 1 min; followed by 35 cycles of denaturation at 94°C for 30 s, annealing at 55-65°C for 30s and extension at 72°C for 1 min. The cycle was finished with a final extension at 72°C for 10 min. The PCR products were assessed on 2% (w/v) agarose gels and visualized by RedSafe (iNtRON Biotechnology, Korea) staining. The amplicon sizes were estimated by comparing against a 50 bp molecular weight ladder (Bioline, Australia). Amplification efficiencies and size polymorphisms were validated in breeding population; parental genotypes ('Sunrise Solo' and 'RB2') and ten segregating progenies from the cross of 'Sunrise Solo' and 'RB2'.

RESULTS AND DISCUSSION

Sequence mapping and identification of polymorphic SSRs in 'Sunrise Solo' and 'RB2' genomes

In total, 30.2 Gb and 32.4 Gb of short reads were generated for 'Sunrise Solo' and 'RB2', respectively, which represented approximately 80-85x coverage of each genome (372 Mb). Over 95% of the reads were successfully mapped to the papaya reference sequence, indicating that whole genome shotgun sequencing was suitable for reconstructing these papaya genomes. Subsequently, 236 Mb of 'Sunrise Solo' and 239 Mb of 'RB2' genomic sequences were produced. There were 173,732 and 173,921 microsatellite fragments identified from 'Sunrise Solo' and 'RB2' genome assemblies, respectively. Both genomes contained a similar density of repeat units (636 SSR/Mb) and this was similar to that detected in a previous papaya SSR study (Vidal et al., 2014), who reported a SSR density of 656 SSR/Mb in the papaya reference genome.

In terms of SSR distribution, the most common SSR types detected in 'Sunrise Solo' and 'RB2' assemblies were mononucleotides (52.6%), followed by dinucleotides (37.4%) and

trinucleotides (7.8%) (Figure1). The frequency of SSRs reduced as repeat units increased in length. Similar results were reported by Wang et al. (2008) and Vidal et al. (2014). Moreover, differential distribution of SSR motif was observed among ‘Sunrise Solo and ‘RB2’ genomes. The majority of dinucleotide SSR fragments were AT/TA units, while fragments containing CG and GC motifs were rare. In addition, AAT/TAA units were abundant compared to other types of trinucleotide motifs. Overall, there was a high AT content in the papaya genomes assessed, which was again consistent with the papaya reference genome (Wang et al., 2008).

For the efficient development of polymorphic SSR markers, sequence variants between the two genomes were analysed *in silico* (Table 1). There were 1,127 polymorphic SSR-containing sequences found with an average motif length of 17 bp. Among these, mononucleotide motifs had the highest percentage of repeat units (55.9%). The second largest group was dinucleotide motifs (39.7%), followed by trinucleotide types (3.4%). Similarly, mononucleotide and dinucleotide SSR motifs were reported to occur most frequently with the highest level of variation in other fruit crops such as peach (Sosinski et al., 2000), tomato (He et al., 2003) and apricot (Sánchez-Pérez et al., 2005).

Figure 1: Occurrence of SSRs in the genomes of ‘Sunrise Solo’ and ‘RB2,’ and frequency distribution of dinucleotide and trinucleotide repeat motifs.

Table1: Distribution of putative polymorphic SSRs identified between ‘Sunrise Solo’ and ‘RB2’

Repeat unit	Count	%	Average motif length	Number of repeats							
				4	5	6	7	8	9	10	>10
Mononucleotide	630	55.9	16	0	0	0	0	0	0	144	486
Dinucleotide	447	39.7	22	0	0	41	25	22	17	7	335
Trinucleotide	38	3.4	18	0	18	10	4	5	0	1	38
Tetranucleotide	9	0.8	22	0	7	0	1	1	0	0	0
Pentanucleotide	3	0.3	27	0	2	1	0	0	0	0	0
Hexanucleotide	0	0.0	0	0	0	0	0	0	0	0	0
Total	1127	100	17.33	0	27	52	30	28	17	152	859

Taking advantage of whole genome sequencing data and an increasingly larger number of functional genes being reported in the literature, the polymorphic SSR-containing sequences were further investigated to predict association with functional genes and CDS regions. All identified coding regions were assigned at least one functional annotation from gene ontology terms. A total of 30.1% (478 annotations) of the polymorphic SSR-containing sequences were subsequently associated with at least one GO term. These encompassed all three major functional categories of biological processes, cellular components and molecular functions (Figure 2). However, the majority of polymorphic SSRs were located in non-coding regions of the genomes, as reported in previous studies (Sun et al. 2013 and Vidal et al., 2014).

Figure 2: GO term distribution of CDS containing SSR polymorphic sequences.

Validation of polymorphic SSR markers

Of the fifty SSR loci assessed for ability to amplify within the parental genotypes and progeny, all (100%) produced single PCR products (Table2, Figure 3). All but six primer pairs produced amplicons of the predicted sizes, which produced larger than expected sizes. This was potentially a result of incomplete sequence assembly within the regions between the primer binding sites. A total of thirty-eight primers (76%) yielded polymorphic amplicons between ‘Sunrise Solo’ and ‘RB2’. These polymorphic SSR loci generated 84 alleles with an average of 2.33 alleles/locus. The result was slightly higher than detected in the study of Oliveira et al. (2012), which demonstrated an average of 2.19 alleles/locus among 36 papaya inbred lines. In this study, heterozygosity was found in 5 loci: CT110 CT241 CT262 CT32-2 and CT9395, as reflected by multiple bands present in each parental genotype. By combining whole genome sequencing, functional annotation and *in silico* polymorphism analysis, this study significantly increased the efficiency of polymorphic marker development for the target genotypes. The novel polymorphic SSR loci will be applied to the future mapping of QTL harbouring desirable traits, potential candidate gene identification and subsequent marker-assisted breeding strategies using these genotypes.

Table2: Primer sequences, SSR motifs and degree of polymorphism of the 50 novel SSR loci.

Primer name		Sequence 5' to 3'	T ^m	Motif	# of allele	Size range (bp)	Polymorphism	GO Description
CT0	F	TCACTGATTCAATTTCCGGCA	67.8	(AT)9(TGTA)5	2	949-1034	polymorphism	probable-phosphate phosphatase C
	R	ATATGTGGTGTGGGGGCA	67.4					
CT104	F	AGGCAGAAATGGGAGACACG	66.8	(A)15(T)12(A)12	2	318-361	polymorphism	hydroxymethylglutaryl-mitochondrial-like
	R	GACCCCAATCACGAAACGATA	66					
CT110	F	AGTTTCGAGTCAAATGGAGGG	64.5	(C)15	3	387-463	polymorphism	nicotinamide 2-like
	R	TGGGCTTCATCTTGGTCCAT	67					
CT1140	F	TATATCAGCACCCACCCCA	67.5	(GA)23	2	832-867	polymorphism	E3 ubiquitin-ligase CHFR isoform
	R	CAGACAGTGCTCACCCCTTT	64.7					
CT126	F	AGCTTCTCATAACCTTAGCCAA	61.2	(CAG)11	2	399-616	polymorphism	growth-regulating factor 3 isoform X1
	R	CCGATTGCAGACTTTGGACC	67.1					
CT149	F	ATGCTTCAACGGCCATAGTT	63.7	(AG)26	2	480-512	polymorphism	Transcription factor chloroplastic mitochondrial
	R	AGGTAGCAACAAGTTTCATCAG	63.6					
CT1492	F	ATGCTTCAACGGCCATAGTT	63.7	(AG)12	2	480-518	polymorphism	Pentatricopeptide repeat superfamily
	R	AGGTAGCAACAAGTTTCATCAG	63.6					
CT150	F	GGATTTTGATACCGTTCATACCC	64.3	(TA)8(AAT)6	2	431-525	polymorphism	BUD31 homolog2
	R	CACTTTCCCAAACGAAGCCTAA	66.3					
CT151	F	ATGTGTCTACTTCGGGCTCC	63.5	(TTAT)5	2	437-472	polymorphism	chloride channel CLC-d isoform X2
	R	AGAACAAAACACCAACCGT	63.6					
CT152	F	TGGCGCTTTGTTAAATTGGC	67.1	(T)17(A)11	2	276-345	polymorphism	E3 ubiquitin-ligase ATL4-like
	R	ACATGGTGAAGAAGCGAATAAC	61.3					
CT161	F	CCCTTTTCATTTCCGTTTCGTA	66.8	(AC)7(AT)7	2	359-406	polymorphism	DNA damage binding 1
	R	TTTACTGCGCCTACTAACGA	61.8					
CT162	F	CCAAGAAAGAAGTTTGTATCCAAG	57.5	(T)18	1	800-842	-	Retrotransposon Ty3
	R	TCCCAACCAAGTACCGCTTC	60					
CT1731	F	ACCGTAAATGAGAGGTTGTGAT	57.4	(CT)25(AT)8	2	201-232	polymorphism	SH2 domainB, 40S ribosomal S28
	R	GTGGAGTAGGTTTGTCCCC	57.8					
CT184	F	GCTTGCCCTTTTATACACACA	58.5	(AT)8(TATG)5	2	844-1000	polymorphism	pentatricopeptide repeat-containing, UDP-glucuronate:xylan alpha-glucuronosyltransferase2
	R	TCTGTTCGTATGCCACCTTTG	58.5					

Table2: Primer sequences, SSR motifs and degree of polymorphism of the 50 novel SSR loci (continue)

Primer name		Sequence 5' to 3'	T ^m	Motif	# of allele	Size range (bp)	Polymorphism	GO Description
CT1939	F	GCTGTTGCTAGCCCAAGG	64	(TTTAA)5	2	287-368	polymorphism	Retrotransposon Ty3-gypsy subclass
	R	TGCATGATGTTCTTCTAACTTGCA	65.8					
CT194	F	TTGCAGCCCGTTAGTTTGG	59.3	(AT)16	1	471-501	-	Peroxisome biogenesis1
	R	ACAACGTCCTCTGCTCATCA	59					
CT204	F	ACGTATCCAGACAGTCCCAA	64	(AAC)5	2	170-203	polymorphism	Two-component response regular ARR5
	R	ATGGTCCCTGTACTCCTCCT	65.2					
CT209	F	GGTTTCATCGGAGGCAGTTTC	58.9	(TA)13	1	216-264	-	Ribonuclease H
	R	CACCGAACCAAAACCTTAGC	59.1					
CT2092	F	TTGTGCTTGAAAATGCCCA	68.8	(CAT)5	2	435-477	polymorphism	Expansin-A6 -like
	R	AGTTCTTCCGTTGGTGGGTA	63.7					
CT215	F	TCACTTAATCCCTTGCCAAAGT	63.9	(TAT)6	2	625-709	polymorphism	Hypothetical protienT01_1515
	R	AAAATGCATGACACCTGTACAAA	63.4					
CT2-2	F	ACTCTTGGAAGTCATATAGGCAA	61	(T)10(AT)10	2	500-718	polymorphism	Zinc transporter 8-like, DNA-RNA polymerase superfamily
	R	CGTTTCTAGTTTCTGCATGTG	62.7					
CT2226	F	TGTCATTTATGTACAAATTGCAC	66.6	(TA)15	2	452-495		Transcription factor MYB44
	R	TGGACAACAACGATTTTCAACT	63.6					
CT23	F	ACTCCGCAGTGTGTAATATTGA	62.6	(AT)13	1	517-630		oliopeptide transporter 1-like
	R	AGAGAGGGGTAAAAGTGGCA	63					
CT232	F	GACATCTCCGAATACAGCGC	65.1	(TG)6(AC)6	2	466-541	polymorphism	Oligopeptide transporter 1 -like
	R	CTTCAACGCCAAACCAGGAA	67.5					
CT241	F	TACCCACCTTTACCTCCGTG	62.5	(TA)17	4	560-603	polymorphism	Cation calcium exchange 1 -like
	R	TGTTGGTAAGATTGAGATTGCAG	61.9					
CT254	F	TTTGCAATTTGTCCCGCTGG	60	(A)20(TA)12	1	805-838	-	Heatshock binding isoform
	R	AGAGCAATCACCTTGGGAGC	60					
CT2544	F	ACTGGCCAAGCAAATCGGTA	60	(CT)8	1	601-692	-	RNA directional DNA polymerase
	R	AGCATCAAGAATCCCCTGCA	59.4					
CT262	F	TGCATGCATGTCCCACCTAT	66.7	(TA)20	3	725-829	polymorphism	Transcription factor bHLH94-like
	R	TCAAGACGCATCTCTGGACG	67					
CT28	F	TCCAGGAATCAACCTTGCT	58.9	(T)17(CA)7	2	233-243	polymorphism	Homeobox-lucine zipper HAT4
	R	ACAACCTTAACATGGCAGAGCA	58.1					
CT280	F	ACTTCGCTCCAGGTCAAGTG	60	(T)23	2	693-794	-	aquaporin TIP1-1
	R	ACTACTGCCAGCCACATTCC	60					
CT28-2	F	TGGGGTCCAGGCTTATTGA	66.7	(AT)17	2	391-453	polymorphism	Nicotinamidase 2-like
	R	ACAGCACCGACCTTTCTTTC	63.6					
CT3-5	F	AGGTCAGTGTGCGCAAGTTA	62.8	(AAT)14	2	282-372	polymorphism	Probable sucrose-phosphate synthase 4
	R	GTTCTCGTGCAACCGATACC	64.8					
CT32	F	GCATTCCCAAAATCTGCTGG	66.9	(AT)16	2	369-409	polymorphism	Ribosome biogenesis protein WDR12, monoacylglycerol lipase ABHD6
	R	GGAGCTACAACAAAGGGCA	64.4					
CT32-2	F	GGTTTGTAGAGAGGAGGGTGA	62.5	(AT)16	3	761-877	polymorphism	RIK isoform
	R	TGGAAGGGTGGTATTTTGATGC	67					
CT36	F	GCTTCCATTATTGCACTTGTGG	66.6	(CT)33	2	393-862	polymorphism	Dihydroxy-acid choloplastic
	R	ATCCCACCAAGCCTCATC	66.6					
CT46-1	F	TGCTTGTGATCCTTTAGCGT	57.5	(AG)12	1	240-252	-	RNA recognition motif-containing isoform1, dna RNA polymerase
	R	GAATCCATCCAAGGTTGCCA	58.1					
CT46-2	F	GGCAGGCATTGTAGTAACCG	64.6	(TA)15	2	551-610	polymorphism	Mitochondrial isoform X1
	R	GGAAGTGGGTGTAAAGGAAACT	63.5					
CT426	F	ATCAAGTGGTGGCGCATGTA	66.8	(AT)9(TA)12	2	694-730	polymorphism	Cyclin family

Table2: Primer sequences, SSR motifs and degree of polymorphism of the 50 novel SSR loci (continue)

Primer name		Sequence 5' to 3'	T ^m	Motif	# of allele	Size range (bp)	Polymorphism	GO Description
CT50	F	AATGGCAGTCTCAGTTACCG	62.2	(A)11(AT)10	2	370-454	polymorphism	Transcriptional factor MYB12, pentatricopeptide repeat-containing mitochondrial
	R	TCAAGTTCGATGTGCTCAGC	64.7					
CT524	F	TAGTTGGGATAGGTGGCACA	63.4	(AG)12	2	246-276	polymorphism	Transcription factor HBP-1a
	R	TGAGTACAGCGGTGAATTACTT	63.3					
CT5483	F	GACAGTAGAGGACCCCT	60	(AT)10(A)11	1	838-883		Intracellular transport USO1-like
	R	GCAGCCGCAATGAGAATAC	59.7					
CT59	F	ATGTGAAAATTGTTGGAGACCC	57.1	(TA)19	2	997-1035	polymorphism	AT-hook motif nuclear-localized 0
	R	TTTGAAGGGTAAACAGCGGC	59					
CT62	F	TTTATCAAACACGCAGCCC	59.1	(TTATT)5	2	290-450	-	calceinurine-like phosphoesterase family
	R	ATAAAAGCCTGACCTCCGCT	59.1					
CT64	F	ACTTGCTTATCAGTGCCGCT	64.2	(AG)25	2	596-679	polymorphism	7-methylguanosine phosphate-specific 5-nucleotidase A, aldoketo reductase family 4
	R	GAACCTTTCTTCTCCGCATGT	63.7					
CT80	F	CAATTCGCTGATGGTGTCTGA	66.6	(TA)12(TA)17	2	223-281	polymorphism	probable WRKY transcription factor 70
	R	CACCCATGTACTGCTGCAAA	65.2					
CT80-2	F	GCTTTGCCACTTGACACCAG	64	(TA)14(AT)13	2	622-800	polymorphism	Uncharacterized protein TCM_019691
	R	TCTCAAGCTTTGCACACCCA	63.3					
CT882	F	AGCTAACCAAGTTTAATACCGCT	60.2	(TA)13	2	348-403	polymorphism	Hexose transporter, sugar transport 13
	R	TTACTTGGTTGTGGGGTCCG	67.2					
CT8821	F	CCGTGCCCCAAGGTGAAAC	58.9	(TA)11	1	288-988	-	Mavicanin
	R	CGGCATCCTTTGCCCAATAA	58.9					
CT9395	F	CGTGACGGTGAGATATCGTGA	66.4	(TA)11	5	948-992	polymorphism	Hypothetical protein L484
	R	GCGACATGTGGAGTGGAGAT	66.2					
CT9835	F	GGACGTGTCTGGAGCATACT	66.2	(AT)13	2	251-529	polymorphism	Hypothetical protein ROM
	R	TATCCAATGTGGCGCACTGT	67.3

Figure 3: An example of polymorphism detected at SSR loci between ‘RB2’ and ‘Sunrise Solo’.

CONCLUSION

A novel set of SSR loci were characterized from the papaya genotypes ‘Sunrise Solo’ and ‘RB2’ using high density genome re-sequencing. Many of these were associated with function and once fully validated, will be further investigated for use in selective breeding strategies.

ACKNOWLEDGEMENTS

This project has been funded by Horticulture Innovation Australia Limited using the papaya industry levy and funds from the Australian Government.

Literature cited

- Afgan, E., Sloggett, C., Goonasekera, N., Makunin, I., Benson, D., Crowe, M., Gladman, S., Kowsar, Y., Pheasant, M., Horst, R. and Lonie, A. (2015). Genomics Virtual Laboratory: A Practical Bioinformatics Workbench for the Cloud. PLoS One 10(10):e0140829. doi: 10.1371/journal.pone.0140829.
- Blankenberg, D., Taylor, J., Schenck, I., He, J., Zhang, Y., Ghent, M., Veeraraghavan, N., Albert, I., Miller, W., Makova, K.D., Hardison, R.C. and Nekrutenko, A. (2007). A framework for collaborative analysis of ENCODE data: making large-scale analyses biologist-friendly. Genome Res 17(6), 960-964.
- Chen, C., Yu, Q., Hou, S., Li, Y., Eustice, M., Skelton, R.L., Veatch, O., Herdes, R., Diebold, L., Saw, J., Feng, Y., Qian, W., Bynum, L., Wang, L., Moore, P.H., Paull, R.E., Alam, M. and Ming, R. (2007). Construction of a sequence-tagged high-density genetic map of papaya for comparative structural and evolutionary genomics in brassicales. Genetics 177(4), 2481-2491.
- Delaporta, S., Wood, J. and Hicks, J. (1983). A plant DNA miniprep. version II. Plant Molecular Biology Reporter 1, 19-21.
- Dillon, N.L., Bally I.S.E., Hucks, L.A., Wright C.L., Innes, D.J. and Dietzgen, R.G. (2013). Implementation of SSR markers in mango breeding in Australia. Acta Horticulturae 992, 259-267.
- Eustice, M., Yu, Q., Lai, C.W., Hou, S., Thimmapuram, J., Liu, L., Alam, M., Moore, P.H., Presting, G.G. and Ming, R. (2008). Development and application of microsatellite markers for genomic analysis of papaya. Tree Genetics & Genomes 4, 333-341.
- FAOSTAT, (2014). Crop production. <<http://faostat.fao.org/site/567/Default.aspx?PageID=567> - ancor>.
- He, C., Poysa, V. and Yu, K. (2003). Development and characterization of simple sequence repeat (SSR) markers and their use in determining relationships among *Lycopersicon esculentum* cultivars. Theoretical and Applied Genetics 106(2), 363-373.
- Hennig, S., Groth, D. and Lehrach, H. (2003). Automated Gene Ontology annotation for anonymous sequence data. Nucleic Acids Research 31(13), 3712-3715.
- Li, H. and Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. In Bioinformatics 25(14), 1754-1760.
- Ming, R., Hou, S., Feng, Y., Yu, Q., Dionne-Laporte, A., Saw, J.H., Senin, P., Wang, W., Ly, B.V., Lewis, K.L.T., Salzberg, S.L., Feng, L., Jones, M.R., Skelton, R.L., Murray, J.E., Chen, C., Qian, W., Shen, J., Du, P., Eustice, M., Tong, E., Tang, H., Lyons, E., Paull, R.E., Michael, T.P., Wall, K., Rice, D.W., Albert, H., Wang, M.L., Zhu, Y.J., Schatz, M., Nagarajan, N., Acob, R.A., Guan, P., Blas, A., Wai, C.M., Ackerman, C.M., Ren, Y., Liu, C., Wang, J., Wang, J., Na, J.K., Shakhov, E.V., Haas, B., Thimmapuram, J., Nelson, D., Wang, X., Bowers, J.E., Gschwend, A.R., Delcher, A.L., Singh, R., Suzuki, J.Y., Tripathi, S., Neupane, K., Wei, H., Irikura, B., Paidi, M., Jiang, N., Zhang, W., Presting, G., Windsor, A., Navajas-Perez,

- R., Torres, M.J., Feltus, F.A., Porter, B., Li, Y., Burroughs, A.M., Luo, M.C., Liu, L., Christopher, D.A., Mount, S.M., Moore, P.H., Sugimura, T., Jiang, J., Schuler, M.A., Friedman, V., Mitchell-Olds, T., Shippen, D.E., dePamphilis, C.W., Palmer, J.D., Freeling, M., Paterson, A.H., Gonsalves, D., Wang, L. and Alam, M. (2008). The draft genome of the transgenic tropical fruit tree papaya (*Carica papaya* Linnaeus), Nature 452(7190), 991-996.
- Oliveira, E.J., Amorim, V.O., Matos, E.L., Costa, J.L., Silva Castellen, M., Pádua, J. and Dantas, J. (2010). Polymorphism of microsatellite markers in papaya (*Carica papaya* L.). Plant molecular biology reporter 28(3), 519-530.
- Oliveira, E.J., Oliveira, G.A.F., Costa, J.L., Oliveira, V.J., Dantas, A.C.V.L., Dantas, J.L.L. and Pádua, J.G. (2012). Genetic diversity and marker-assisted inbreeding in papaya. Scientia Horticulturae 147, 20-28.
- Pinto, F.O., Pereira, M.G., Luz, L.N., Cardozo, D.L., Ramos, H.C.C. and Macedo, C.M.P. (2013). Use of microsatellite markers in molecular analysis of segregating populations of papaya (*Carica papaya* L.) derived from backcrossing. Genet. Mol. Res. 12(3), 2248-2259.
- Sánchez-Pérez, R., Ruiz, D., Dicenta, F., Egea, J. and Martínez-Gómez, P (2005). Application of simple sequence repeat (SSR) markers in apricot breeding: molecular characterization, protection, and genetic relationships. Scientia Horticulturae 103(3), 305-315.
- Sosinski, B., Gannavarapu, M., Hager, L.D., Beck, L.E., King, G.J., Ryder, C.D., Rajapakse, S., Baird, W.V., Ballard, R.E. and Abbott, A.G. (2000). Characterization of microsatellite markers in peach (*Prunus persica* L. Batsch). Theor Appl. Genet 101, 421-428.
- Sun, L., Yang, W. and Zhang, Q. (2013). Genome-wide characterization and linkage mapping of simple sequence repeats in Mei (*Prunus mume* Sieb. et Zucc.). PLoS ONE 8, e59562.
- Vidal, N.M., Grazziotin, A.L., Ramos, H.C., Pereira, M.G. and Venancio, T.M. (2014). Development of a gene-centered ssr atlas as a resource for papaya (*Carica papaya*) marker-assisted selection and population genetic studies. PLoS One. 9(11), e112654. doi: 10.1371/journal.pone.0112654
- Viruel, M.A., Escribano, P., Barbieri, M., Ferri, M. and Hormaza, J.I. (2005). Fingerprinting, embryo type and geographic differentiation in mango (*Mangifera indica* L.) Anacardiaceae with microsatellites. Mol. Breed. 15, 383-393.
- Walls, R.L., Athreya, B., Cooper, L., Elser, J., Gandolfo, M.A. and Jaiswal, P. (2012). Ontologies as integrative tools for plant science. Am J Bot 99, 1263-1275.
- Wang, X., Lu, P. and Luo, Z. (2013). GMATo: A novel tool for the identification and analysis of microsatellites in large genomes. Bioinformatics 9(10), 541-544.
- Wang, J., Chen, C., Na, J.K., Qingyi, Y., Hou, S. (2008). Genome-wide comparative analyses of microsatellites in papaya. Tropical Plant Biology 14, 278-292.
- Zhang, R.P., Wu, J., Li, X.G., Khan, M.A., Chen, H., Korban, S.S. and Zhang, S.L. (2013) An AFLP, SRAP, and SSR genetic linkage map and identification of QTLs for fruit traits in pear (*Pyrus* L.). Plant Mol. Biol. Rep. 31, 678-687.