Website : www.asthafoundation.in
Frontiers in Crop Improvement
Vol 12 (1) : 28-32 (January-2024) Print ISSN : 2393-8234 Online ISSN : 2454-6011
Astha Foundation, Meerut (U.P.) India

Development and Validation of Microsatellite Markers for Coconut Genotypes obtained by Partial Genome Assembly

Nidhi Savaliya, Rukam S. Tomar and Shital Padhiyar
Department of Biotechnology, Junagadh Agricultural University, Junagadh, Gujarat

Abstract

An Ion S5 Next Generation Sequencer (NGS) was used to sequence the 400bp DNA library of coconut genotypes to obtain a draft genome and for mining of microsatellite markers. The de novo assembly yielded assembled reads of 779 mbp for dwarf genotypes and 872 mbp for tall genotypes. A total of 15 SSRs were identified and examined. A total of 38,782 SSRs were identified from 16,51,372 contigs examined. The number of sequences containing SSR were 35,682 and the number of sequences containing more than 1 SSRs were 2,867. Six SSR markers were validated among the six genotypes of coconut. The average percentage polymorphism of markers was 41.60% with average Polymorphism Information Content (PIC) of 0.23 and average SSR primer index (SPI) of 0.47. These microsatellite markers can be used for linkage mapping and genes/QTLs tagging for genetic improvement of target traits in coconut and related crops.

Key words *: Genome, coconut, SSR.*

Introduction

Belonging of the Arecaceae family, the coconut palm (*Cocos nucifera* L.) is a tropical tree with a 2n=32 ploidy. It is said to have come from the Bismarck Archipelago, New Guinea, the Malayan Peninsula, and the Bismarck Archipelago, from whence it spread across the tropical regions. Its present distribution includes North, Central, and South America, West and East Africa, Southeast Asia, the Pacific Islands, and other regions (1). Nowadays, coconuts constitute a staple meal in tropical parts of the world. Coconut water has a long history of religious associations since it is a clean and sterile beverage. Throughout Asia, and particularly in India, tender, or immature, coconuts are presented as ceremonial gifts and used as cleansing implements at customary ceremonies.

The market for items made from coconuts is expanding annually, but non-food byproducts with or without processing can also be profitable (2). Survey organizations in India, New England, Indonesia, the Philippines, Thailand, and Latin America have been successful in launching new ventures that apply improved or novel techniques to capitalize on coconut products that aren't used as food due to growing social demands on the global market to replace highly polluting oil-derived products-like plastics and diesel, among otherswith environmentally friendly substitutes.

One of the most widely grown palm tree species worldwide, the coconut tree is also known as the "tree of life" or "the tree of a thousand uses" (2). You may use the surface area of your palm. Its trunk and leaves are widely utilized in building materials for fences and other structures, as well as in handicrafts (1). The fruit's fibre mesocarp, or husk, is used to make rope, compressed wood, carpets, geotextiles, inert, sterile support medium for plants, dietary fiber, and cushioning for seats in cars and trains (3). The portions of the fruit, such the coconut kernel and tender coconut water, have several medicinal properties, such as antibacterial, antifungal, antiviral, antiparasitic, antidermatophytic, antioxidant, hypoglycemic, hepatoprotective, and immunostimulant functioning. People eat coconuts because they are a food source in tropical countries and because their water and kernels contain microminerals and nutrients that are vital to human health. Because of this, the coconut palm is regarded in Indian scriptures as "Kalpavriksha" (the tree that gives everything), symbolizing the phenomenon associated with its application in the promotion of health and the avoidance of sickness.

One of the numerous purposes for which coconut trees are grown is for their nutraceutical advantages. Tender coconut water, copra, coconut oil, raw kernels, coconut cake, coconut toddy, wood-based and coconut shell goods, coconut leaves, and coir pith are among the several coconut products. In the traditional coconut farming regions, people make use of every part of the coconut in their everyday life. It is the only source of several natural compounds that may be utilized to make industrial goods and drugs that treat a range of illnesses. Between 65% and 75% of the dried copra kernel, which is mostly utilized for oil extraction, is oil.The whole spathe is tapped for toddy, which is subsequently made into jaggery, vinegar, and sugar. Most often, the kernel (wet meat) is used to make curries, chutneys, toffee, desserts, and other foods (4).

Received : November-2023; Revised : November-2023; Accepted : December-2023

When extracted from young fruit, the solid endosperm has a gelatinous texture that makes it ideal for use as a functional meal or in dessert recipes. Whether grown industrially or artisanally, coconut fruit is a major source of income for coconut farmers in producing nations (2). TCW, or the liquid endosperm of the tender coconut, is a great natural soft drink option. There are 17.4 calories in every 100 grams. The vitamins B found in coconut water include nicotinic acid B3, pantothenic acid B5, biotin, riboflavin B2, folic acid, and trace amounts of thiamine B1 and pyridoxine B6 (5).

Coconut water contains sugars and sugar alcohols as well as free amino acids, vitamin C, folic acid, enzymes (acid phosphatase, catalase, dehydrogenase, diastase, peroxidase, and RNA polymerases), phytohormones (auxin, 1, 3-diphenylurea, and cytokinin), and substances that promote growth (6). Coconut sugar, coconut water, and virgin coconut oil have all been enthusiastically added to the growing global market for alternative health foods. Customers in this sector, however, are looking for assurances on the safety of natural goods. Alternative methods of heating coconut water have been used, such as sterilising at extremely high temperatures, heating in a microwave, filtering, and refrigerating, and treating at both low and high temperatures while adding sulfites (7,8).

Microsatellite markers are the most effective in establishing a marker-trait association, and using molecular markers is a very accurate way to produce cultivars with desired traits. Microsatellites are repetitive DNA segments scattered across the genome in noncoding regions between genes or inside genes (introns). These regions are known as short tandem repeats (STR), simple sequence repeats (SSRs), or simple sequence length polymorphisms (SSLPs) (Litt and Luty, 1989). Microsatellites are defined as 2-8 bp repeats in some studies and as 1-6 or even 1-5 bp repeats in others (9,10). Microsatellites are most commonly defined as repeats of mono-, di-, tri-, and tetra nucleotides; however, they can also be defined as repeats of five (penta-) or six (hexa-) nucleotides. Primer's extension or selective hybridization have been used to enrich genomic libraries, which has historically served as the foundation for microsatellite isolation and identification. Searching DNA databases like EST sequences for microsatellite repeats is an alternative strategy. It is now possible to quickly and economically get a significant number of high-quality genome sequences, including hundreds of microsatellite sites, in the genome of a target species because to advancements in genome sequencing techniques like Next Generation Sequencing (NGS). According to (11,12), microsatellite markers can be used for QTL mapping and linkage map generation. Furthermore, the nutritional composition of coconuts varies significantly according on the variety, geographical location, and stage of growth. In this study, the tall and dwarf genotypes of coconut were used to evaluate differences in the chemical composition of the pulp and water. Consequently, the current effort partly sequenced the coconut genome using NGS. The contig sequences created by partial de novo assembly were subsequently mined for microsatellite sites in order to build crop-specific SSR markers. The genotypes of coconuts were then used to test these markers.

Materials and Methods

DNA extraction and partial genome assembly : The draft coconut genome was meticulously sequenced, leveraging DNA sourced from tall and dwarf genotypes generously provided by the College of Horticulture at Junagadh Agricultural University in Junagadh, Gujarat, India. Employing the DNA extraction method outlined by (13), the DNA was extracted from the young and fresh leaves of both tall and dwarf genotypes. Subsequently, a 400 bp genomic DNA fragment library was meticulously amplified through emulsion PCR using the Ion OneTouchTM 2 System. Following this, positive bead recovery was conducted for template enrichment. The sequencing process involved the Ion S5 sequencer, and the 400 bp fragments were sequenced in adherence to the system's recommended guidelines.

Genome Assembly : Contaminated DNA sequences were identified by subjecting the coconut findings to a BLAST search against a database encompassing potential contamination sources such as bacteria, viruses, and fungi. Any sequences found to be contaminated were promptly excluded from the subsequent analysis. For the removal of adapter sequences introduced during Ion S5 sequencing, the CLC trimmer function with a default limit of 0.05 was employed, utilizing CLC Genomics Workbench 8.2 software from CLC Bio in Aarhus, Denmark. Additionally, the CLC Genomics Workbench software, configured with a maximum stringency of 0.50 for length fraction (LF) and 0.80 for sequence similarity (SIM) between DNA reads, was utilized in the assembly process. To ensure robust assembly, a minimum contig length requirement of 100 bp was enforced.

SSR search and primer design : To identify Simple Sequence Repeats (SSRs), we employed PRIMER 3 software. The criteria for SSR selection included a minimum of six di-nucleotide repeats, five tri-nucleotide repeats, and three repeats for tetra-, penta-, and hexa-nucleotide sequences. Subsequently, primers for the identified SSRs were generated using Batch Primer 3.0 software, adhering to specific parameters: primer lengths ranging from 14 to 23 bases, GC content between

Table-1 : Statistics of partial genome sequencing data of coconut generated with Ion Torrent S5.

	Parameters	Tall Coconut Genotype	Draft Coconut Genotype
1.	No of Contigs (,000)	872	779
2.	Total size (Mb)	369.73	294.47
3.	Maximum length	18,500 bp	14,852bp
4.	Average	424 bp	378bp
5.	N25 (Bases)	1,272 bp	998bp
6.	N50 (Bases)	648 bp	572bp

Table-2 : The statistics of genomic data assessed of SSR identification and frequency distribution of the SSRs in the genomic dataset of coconut.

	Parameters		
1.	Total number of sequences examined	:	1651372
2.	Total size of examined sequences (bp)	:	664207321
3.	Total number of identified SSRs	:	38782
4.	Number of SSR containing sequences	:	35682
5.	Number of sequences containing more than 1 SSR	:	2867
6.	Number of SSRs present in compound formation	:	528

Table-3 : Distribution of di-, tri-, tetra, penta and hexa-nucleotide microsatellites on contigs of Coconut.

	Unit size	Number of SSRs
1.	Di-nucleotide repeats	24,231
2.	Tri-nucleotide repeats	12,234
3.	Tetra-nucleotide repeats	1,782
4.	Penta-nucleotide repeats	457
5.	Hexa-nucleotide repeats	78
	Total	38782

Table-4 : List of SSR primers.

No.	Sequence ID		Tm	GC%	Sequence	Prod size
1.	contig_127.p1	F	54.87	45.00	GAATCTAGTGATGGCAAAGG	206
		R	54.39	47.62	TGACTGGTAAGGAGGATACTG	
2.	contig_165.p1	F	55.04	45.00	CTGATATTGCTGGGGTGTAT	141
		R	55.47	47.62	CTCCTTCTCCAGTCCATTTAC	
3.	contig_270.p1	F	56.18	38.10	ATGGTTCAATAGCAGTTGGAT	186
		R	55.20	42.86	ACAATAGTCTCGAAGCAGTGA	
4.	contig_350.p1	F	58.29	61.11	GCTCCAGGAAAGGAGGTG	211
		R	54.95	47.62	GTCATCTAGCACAGGAATCTG	
5.	contig_549.p1	F	55.08	42.86	GGTACGCGTATGAGTTTTATG	127
		R	55.04	42.86	ACACACTATCAAGCAGAGGAA	
6.	contig_6131.p1	F	55.35	47.62	TATTACACTCACCTCCTCACG	157
		R	55.56	42.11	AGCCGTTCTTTATTTCACG	
7.	contig_7348.p1	F	54.60	47.37	CAGTCCCGATCCTTTAATC	143
		R	55.70	38.10	CATCAATGTGATGGAGATCAT	
8.	contig_767.p2	F	56.84	55.56	CAGGTTCAGTCCGTGGAT	150
		R	55.24	50.00	TACTGGCTGGACAAGGTATC	
9.	contig_1304.p1	F	55.56	38.10	CATTAGCTTGTTCTCCATTCA	179
		R	54.52	42.86	CCAGATTCAGAAGGTCTACAA	
10.	contig_5533.p1	F	55.61	52.63	GAAGAGCTCATCACCCAAC	171
		R	55.00	42.86	TAGAAGAAGGGGAAGAAGAGA	
11.	contig_6864.p1	F	54.61	42.86	TAAAGTCTCTTGAGGTGGTTG	154
		R	56.07	38.10	CAACTCCCCATATTCATTTTC	
12.	contig_6926.p1	F	55.05	38.10	TTTATCTTCTCCCTTCAATCC	213
		R	56.51	61.11	GTCACCGGCAGGACTTAC	
13.	contig_416.p2	F	54.76	45.45	CATCATGTCTATGGTAGTACGG	143
		R	54.50	50.00	GTGACCTTCTCCATGAACTC	
14.	contig_5920.p1	F	55.54	42.86	CCTGTCGCTCTATCTTCTTTT	172
		R	54.65	33.33	GTTGAAGCATAAGAAATTCCA	
15.	contig_6131.p1	F	55.35	47.62	TATTACACTCACCTCCTCACG	157
		R	55.56	42.11	AGCCGTTCTTTATTTCACG	

Table-5 : Size, number of amplified bands, percent polymorphism and PIC obtained by six SSR primers in the six genotypes of coconut.

Sr. No.	SSR primers	Allele/Band Size	Total no. of allele/ band	Polymorphic band			Monomorphic Band	% polymorphism	PIC	SPI
				S	U	T				
1.	contig_767.p1	235	1	0	0	0	1	0.00	0	0
2.	contig_250.pl	200-250	2	0	1	1	1	50.00	0.24	0.48
3.	contig_1340.pl	250-300	2	2	0	2	0	100.00	0.5	1
4.	contig_416.pl	300-350	2	0	1	1	1	50.00	0.24	0.48
5.	contig_6131.pl	200-250	2	1	0	1	1	50.00	0.44	0.88
6.	contig_5920.p1	248	1	0	0	0	1	0.00	0	0
	Total	10	3	2		5	5	-	-	-
	Average	-		-		0.83	0.83	41.60	0.23	0.47

45% and 60%, an annealing temperature set between 50°C and 60°C, and a desired PCR product size of 200 bp and above.

Screening and Assessment of SSR : Random primers for the genome were sourced from Merck, India. DNA isolation from six coconut genotypes namely Green Draft, Orange Draft, Yellow Draft, West coast Tall, Cochin China Tall, and Andaman Ordinary Tall was carried out using the DNeasy Plant Mini Kit (Qiagen, Valencia, CA). The polymerase chain reaction (PCR) was executed with the following components: one microliter (50 ng) of template DNA, one microliter each of forward and reverse primers, two microliters of 10x Taq buffer + $MgCl_2$ (15 mM), two microliters of dNTP (2 mM), four microliters of Taq polymerase (Promega, 5U L^{-1}), and two microliters of fresh distilled water. The PCR reaction took place in an Applied Biosystems Veriti heat cycler, involving 30 seconds of annealing at 58–59°C, an initial denaturation step at 94°C for 30 seconds, and a final extension at 72°C for seven minutes. Subsequently, the amplified PCR products for each primer set underwent electrophoresis using a 100 bp DNA ladder from Qiagen in Valencia, California, on a 3% agarose gel at 120 V for approximately 1.5 to 2.0 hours. Gel imaging was performed using the Gel Documentation system from Syngene, an Indian company.

Results and Discussion

Genome Assembly : All readings meeting the quality filtration criterion (mean quality score >=20) were utilized for assembly preparation, as detailed in the materials and techniques section, employing the CLC Workbench. Subsequently, the data from Ion S5 runs were combined and utilized for assembly using the CLC Workbench V8.2 de novo assembler. The draft genome of the Andaman Ordinary Tall coconut genotype, representing tall genotypes, was generated, comprising 369.73 million base pairs. In contrast, the draft genome of the Yellow Dwarf genotype, representative of dwarf genotypes, consisted of 294.47 million base pairs.

For the Andaman Ordinary Tall coconut genotype, the contig lengths varied from a minimum of 100 bp to a maximum of 18,500 bp, with an average contig length of 424 bp. The N25 and N50 contig sizes, representing the length at which 25% and 50% of the total genome is covered, were determined as 1,272 bp and 648 bp, respectively (Table-1).

In the case of the dwarf genotype, the contig lengths ranged from 100 bp to a maximum of 14,852 bp, with an average contig length of 378 bp. The N25 and N50 contig sizes were found to be 998 bp and 572 bp, respectively.

SSR Mining : A comprehensive analysis using the MISA tool on a total of 1,651,372 contigs resulted in the identification of 38,782 Simple Sequence Repeats (SSRs) (Table-2). Among these, 2,867 sequences contained more than one SSR, and 528 SSRs were observed in compound form. The identified SSRs comprised five types of repeated motifs: dinucleotide, trinucleotide, tetranucleotide, pentanucleotide, and hexanucleotide. Notably, dinucleotide and trinucleotide SSRs were the most prevalent. The distribution of SSRs across di-, tri-, tetra-, penta-, and hexa-nucleotide motifs revealed counts of 24,231, 12,234, 1,782, 457, and 78, respectively (Table-3).

Primer Design and Validation : A set of 15 primers were designed with the product size varying between 200 to 300bp (Table-4). The primers were designed from the different contigs assembled from the reads. The validation of these primers was carried out on six genotypes of coconut*viz.,*Green Dwarf, Orange Dwarf, Yellow Dwarf, West Coast Tall, Cochin China Talland Andaman Ordinary Tall.A total of 15SSR primers were synthesized from contigs in which six primers were selected. A total of 26 bands were amplified. The SSR marker from the Contig_767.pl produced maximum number of 2 bands, while, primer from contig_250.pl, Contig_1340.pl,

Contig_416.pl and contig_6131.pl produced only 1 band.All six primers gave an amplification at the expected base pair with a total number of 10 amplified bands. Based on the sequence data produced by the *de novo* assembly technique, expected product sizes for each microsatellite marker were calculated. We verified that the polymorphic loci's size ranges in the size of the expected product. Out of the total number of bands amplified by the 6 SSR primers, 5 were polymorphic and 5 were monomorphic with an average polymorphism of 0.83 (Table 5). The average percentage polymorphism was 41.60% with average polymorphism information content (PIC) of 0.23 and average SSR primer index (SPI) of 0.47.

The capacity to discover and create microsatellite markers for molecular breeding, is undergoing a true revolution. A new paradigm of microsatellite development has been put to the test by research teams with access to an NGS facility, and continual decline in pricesfor obtaining next-generation sequencing data has provided easy access for obtaining genomic sequences. Shotgun pyrosequencing of DNA or enriched libraries were primarily employed in the majority of the initial articles published reporting the use of next-generation sequencing technologies for the production of microsatellite markers. Similar findings of (14) effectively synthesized, optimized, and amplified 131 new SSR markers in coconut utilizing genotypes 'CATD,' 'LAGT,' 'WAT,' and (LAGT WAT). Among the 131 SSR markers tested, 113 were polymorphic in the coconut genotypes. (8) created a total of 7,139 distinct SSR markers with the intention of being used as a resource in marker-based breeding. Furthermore, by matching the Hainan Tall (HAT) WGS reads to the non-repetitive sections of the completed CATD genome and detected 58,503 variations in coconut.

References

1. Prades A., Dornier M., Diop N. and Pain J.P. (2012). Coconut water, composition and properties: a review. *Fruits,* 67: 87–107.
2. Angeles J.G., Lado J.P., Pascual E.D., Cueto C.A., Laurena A.C. and Laude R.P. (2018). Towards the understanding of important coconut endosperm phenotypes: is there an epigenetic control? *Agronomy*, 8: 225. https://doi.org/10.3390/agronomy8100225.
3. Rencoret J., Ralph J., Marques G., Gutierrez A., Martinez A.T. and del Río J.C. (2013). Structural characterization of lignin isolated from coconut (*Cocos nucifera L.*) coir fibers. *Journal of Agricultural and Food Chemistry,* 61: 2434–2445.
4. NMCE. Report on copra. National Multi-commodity Exchange of India Limited; (2007), *pp*. 1-14.
5. United States Department of Agriculture (USDA). National nutrient database for standard reference, Nuts, coconut water, (2008). [Online]. Available from: http://www.nal.usda.gov/fnic/foodcomp/ cgi-bin/list_nut_edit.pl/. [Accessed on December 8, 2009].
6. Yong J.W.H., Ge L., Ng Y.F. and Tan N. (2009). The chemical composition and biological properties of coconut (*Cocos nucifera* L.) water. *Molecules,* 14: 5144–5164.
6. Litt M. and Luty J.A. (1989). A hypervariable microsatellite revealed by in vitro amplification of a dinucleotide repeat within the cardiac muscle actin gene. *American Journal of Human Genetics,* 44: 397-401.
7. Sucupira N.R., Alves F.E.G., Silva L.M.A., de Brito E.S., Wurlitzer N.J. and Sousa P.H.M. (2017). NMR spectroscopy and chemometrics to evaluate different processing of coconut water. *Food Chemistry,* 216: 217–224.
8. Nirmala Maurya, Mukesh Dixit and Subrata Pani (2023). Assessment of macrobenthic invertebrates of a central Indian water body (Jobat Dam) with reference to its water quality. *Progressive Research : An International Journal*, 18(1): 67-76.
9. Schlotterer C. (1998). Microsatellites. In: *Molecular Genetic Analysis of Populations: A Practical Approach* (Ed. A.R. Hoelzel). IRL Press, Oxford. 5: 237-261.
10. Nagesh CR, Tamil S, Dutta M, Indrajit WS, Dinesh KR, Sushmitha J, Durgeshwari G, Shukla A, Laishram R, Bansal N, Rama PG, Goswamy S, Kumar RR, Bhardwaj C, Tyagi A, & Vinutha T. (2023). Phytoremediation: green to clean environmental heavy metal pollution. Agrisustain-an International Journal, 01(01), 16–20.
11. Tomar R.S., Parakhia M.V., Thakkar J.R., Rathod V.M., Padhiyar S.M., Thummar V.D., *et al.* (2016). Development of linkage map and identification of QTLs responsible for Fusarium wilt resistance in castor (*Ricinus communis* L.). *Research journal of biotechnology*, 11: 67–73.
12. Tomar R.S., Parakhia M., Rathod V., Thakkar J., Padhiyar S., Thummar V., Dalal H., Kothari V., Kheni J. and Dhingani R. (2017). Molecular mapping and identification of QTLs responsible for charcoal rot resistance in castor (*Ricinus communis* L.*). Industrial Crops and Products,* 95: 184-190.
13. Desai H., Hamid R., Ghorbanzadeh Z., Bhut N., Padhiyar S.M., Kheni J. and Tomar R.S. (2021). Genic microsatellite marker characterization and development in little millet (*Panicumsumatrense*) using transcriptome sequencing. *Scientific Reports,* 11: 206-220.
14. Caro R.E.S., Cagayan J., Gardoce R.R., Manohar A.N.C., Canama-Salinas A.O., Rivera R.L., Lantican D.V., Galvez H.F. and Reano C.E. (2022). Mining and validation of novel simple sequence repeat (SSR) markers derived from coconut (*Cocos nucifera* L.) genome assembly. *Journal of Genetic Engineering and Biotechnology*, 20: 71.