Functional & Integrative Genomics (2022) 22:1243–1251
https://doi.org/10.1007/s10142-022-00911-2

RESEARCH

Development and characterization of non-coding RNA-derived simple sequence repeat markers in coconut (*Cocos nucifera* L.)

A. A. Sabana[1] · Ginny Antony[1] · M. K. Rajesh[2] · K. P. Gangaraj[2] · V. Niral[2] · R. Sudha[2] · B. A. Jerard[3]

Received: 17 August 2022 / Revised: 17 October 2022 / Accepted: 18 October 2022 / Published online: 1 November 2022

Abstract

Non-coding RNA (ncRNA)-based SSR markers are highly useful in molecular breeding as ncRNAs play a significant role in gene regulation. In the present study, for the first time in coconut, we have identified 597 ncRNA-derived SSR markers, including 509 long non-coding RNASSRs (lncRNASSRs) and 88 micro RNASSRs (miRNASSRs). Of these, 20 primers (10 each from lncRNA-SSR and miRNA-SSR) were selected, screened on 6 coconut accessions, and 50% produced polymorphic fragments. These 10 polymorphic primers were used for genotyping 96 palms of 16 coconut accessions, comprising eight tall and dwarf accessions each. The number of alleles ranged from 2 to 9 per SSR marker, with an average of 4.6 alleles per locus. The average heterozygosity and Shannon index were 0.5 and 1.1, respectively, suggesting that ncRNA-SSRs show high polymorphism level. Distance-based cluster analyses revealed that all the tall and dwarf accessions were differentiated and grouped in different clusters. The study demonstrates the usefulness of ncRNA-based SSR markers for assessing genetic diversity and genetic improvement in coconut.

Keywords Coconut · Genetic diversity · lncRNA · miRNA · Non-coding RNA · Polymorphic SSRs

Introduction

Molecular markers have been extensively used for various studies, such as plant molecular breeding, population genetics, and QTL mapping (Nadeem et al. 2018). They have become useful tools for detecting polymorphisms among individuals of populations or species. Various types of molecular markers have been used depending on the characteristics of the marker and the needs of the research objectives (Arunachalam and Rajesh 2017; Rajesh et al. 2019; Ramesh et al. 2022). These include random amplified polymorphic DNA (RAPD), restriction fragment length polymorphism (RFLP), amplified fragment length polymorphism (AFLP), simple sequence repeats (SSR), start codon targeted polymorphism (SCoT), and single nucleotide polymorphism (SNP) (Upadhyay et al. 2004; Lebrun et al. 1998; Teulat et al. 2000; Geethanjali et al. 2018; Rajesh et al. 2015, 2021). Among these markers, SSR markers are widely used in many plant species, such as coconut, for diversity studies and fingerprinting owing to their co-dominant nature, high polymorphism and reproducibility, and more information content.

The development and characterization of novel molecular markers are a constant interest in the field of molecular breeding. Two novel non-coding RNA (ncRNA)-based SSR markers have been recently reported, viz., microRNA-based SSRs (miRNA-SSRs) and long non-coding RNA-based SSRs (lncRNA-SSRs). These markers are considered rare and have not been characterized in many plant species (Jaiswal et al. 2020). Non coding RNAs in plants have important roles in the regulation of many biological processes through transcript modification and translational regulation. The ncRNA-based SSR markers are gaining prominence in the molecular breeding of crops such as capsicum (Jaiswal et al. 2020), *Brassica napus* (Summanwar et al. 2020), *Cleistogenes songorica* (Kanzana et al. 2020),

✉ Ginny Antony
ginnyantony@cukerala.ac.in

✉ M. K. Rajesh
rajesh.mk@icar.gov.in

[1] Central University of Kerala, Kasaragod 671320, Kerala, India

[2] ICAR-Central Plantation Crops Research Institute (ICAR-CPCRI), Kasaragod 671124, Kerala, India

[3] Andaman and Nicobar Islands, ICAR-Central Island Agricultural Research Institute (ICAR-CIARI), Port Blair 744101, India

Triticum aestivum (Bhandawat et al. 2020), *Medicago truncatula* (Min et al. 2017), *Oryza sativa* (Mondal and Ganie 2014; Tabkhkar et al. 2020), *Nelumbo nucifera* (Wang et al. 2016), and *Punica granatum* (Patil et al. 2020). Since these markers are based on the sequence variations in the regulatory RNAs, they are useful in functional diversity analysis in cell differentiation and stress tolerance. Many SSR markers responsive to drought, heat, salt stress has been developed in a variety of plants (Mondal and Ganie 2014; Singh et al. 2018; Mehta et al. 2021; Brake et al. 2022).

Many factors like various diseases, low productivity of coconut populations, and the unstable market price of coconut oil and other coconut-based products have resulted in a gradual decline in coconut cultivation in the tropical regions. It is imperative to develop high-yielding hybrids and new coconut varieties using accessions with diverse traits to address these challenges. Molecular markers are best suited for assessing genetic diversity and varietal identification. Multiple studies reported the development of molecular markers in coconut palms and diversity studies using various molecular markers (Lebrun et al. 1998; Teulat et al. 2000; Manimekalai et al. 2007; Rajesh et al. 2008, 2015; Riangwong et al. 2020; Preethi et al. 2020; Caro et al. 2022). Whole-genome sequence of coconut has also been deciphered recently by three independent groups (Xiao et al. 2017; Lantican et al. 2019; Rajesh et al. 2020). Since ncRNA-based molecular markers in coconut have not yet been reported, in the present study, we have developed ncRNA-based, highly reproducible co-dominant SSR markers from lncRNAs and miRNAs. The markers developed were utilized to assess the genetic diversity of a panel of world-wide coconut collections, comprising of tall and dwarf accessions.

Methods

SSR mining from lncRNAs and miRNAs and primer designing

The pipeline used in this study to identify lncRNA and miRNA-derived SSR markers has been provided in Fig. 1. Briefly, a total of 110 mature miRNAs generated earlier by small RNA sequencing in coconut (Sabana et al. 2020) were used for SSR mining. Pre-miRNA sequences were identified from the coconut genome sequence using a pre-miRNA prediction tool in a small RNA workbench (http://srna-workbench.cmp.uea.ac.uk/). Then, their corresponding sequences, along with 500 bp from the upstream and downstream flanking regions, were used. These 1000 bp sequences and corresponding pre-miRNA sequences were considered primary miRNAs (pri-miRNA). Further, these long pri-miRNA sequences were scanned for mining of SSR motifs using the MIcroSAtellite identification tool (MISA,

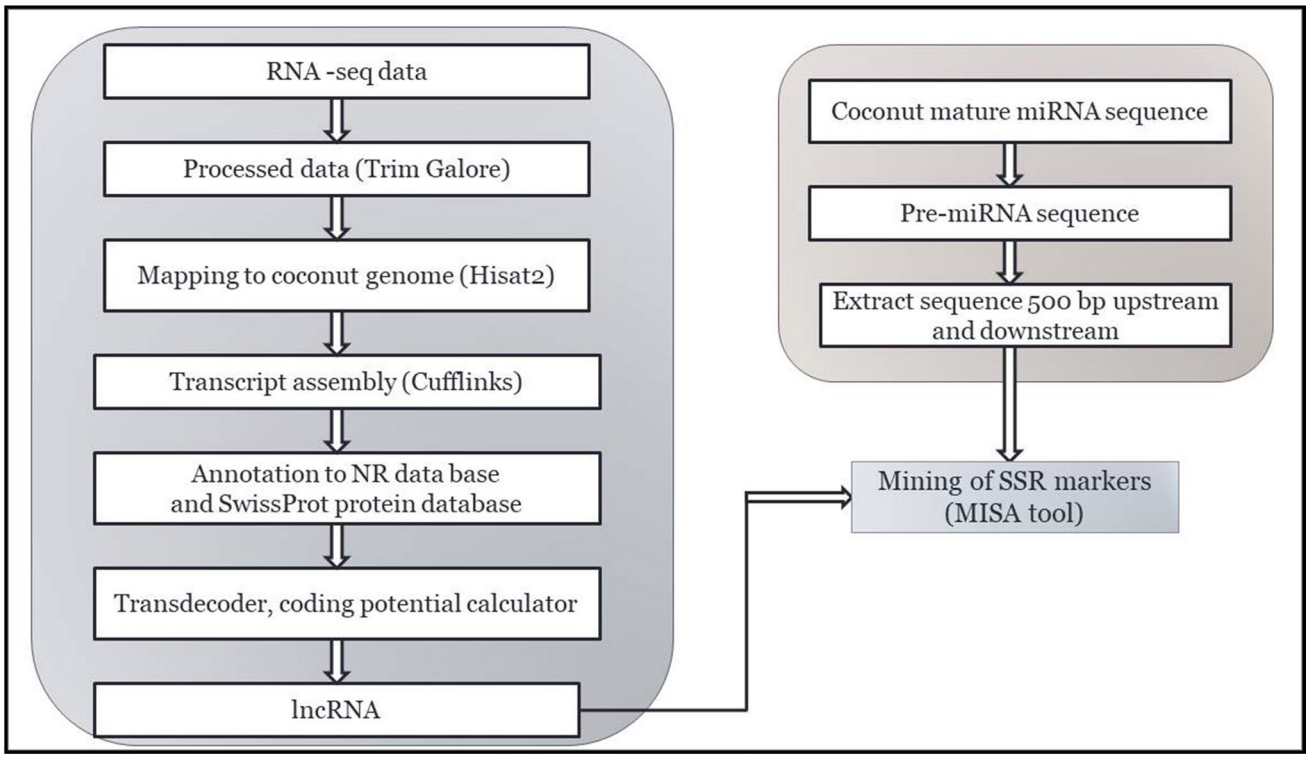

Fig. 1 Schematic diagram shows the bioinformatics pipeline used for the study

http://pgrc.ipk-gatersleben.de/misa/). A total of 5110 coconut lncRNA sequences were used for mining miRNA-SSR using the MISA search tool. The minimum repeat length was set at 20 for mono-nucleotides and 6 for di-, 5 for tri- and tetra- and 3 for penta- and hexa-nucleotides. The maximum number of bases interrupting between two SSRs in a compound microsatellite type was 100.

Further, markers with AT/AT and TA/TA repeat motifs were excluded in the selection of markers manually, according to the procedure given by Caro et al. (2022). SSR primer pairs from flanking sequences were designed using BatchPrimer3 software (https://wheat.pw.usda.gov/demos/BatchPrimer3/). The default parameters were set as follows: primer size ranged from 18 to 23 bp with 21 as optimum; GC content 40–60% with 50% as optimum; melting temperature (Tm) ranged from 57 to 60 °C with 50 °C as optimum temperature and product size ranged from 100 to 300 bp.

For the prediction of chromosome linkage groups, ncRNA-SSR containing sequences were aligned with the reference genome (PRJNA374600) (Yang et al. 2021) using BLASTn (e value:$1 \times e^{-5}$). The alignment locations were manually curated (Supplementary Table 1).

Plant materials and extraction of DNA

Coconut leaf samples were collected from National Coconut Genebank, ICAR-CPCRI, Kasaragod, Kerala State and International Genebank for South East Asia and the Middle East (ICG-SAME), Kidu, Karnataka State, India. A total of 96 palms belonging to 16 coconut accessions (eight tall and dwarf accessions each; six palms per accession) were used for the diversity study (Table 1). Genomic DNA was extracted from spindle leaves using the modified SDS method (Rajesh et al. 2013). In brief, one gram of spindle leaf portion was cut into small pieces, frozen in a pre-chilled mortar with liquid nitrogen, and ground to a fine powder. About 0.25 g polyvinylpolypyrrolidone (PVPP, MW 400,000) was added to the ground leaf samples, mixed thoroughly, and transferred to a centrifuge tube containing 4.5 ml preheated extraction buffer. Then, 0.5 ml of 10% SDS and 50 µl of β-mercaptoethanol were added to the mix, and the mixture was incubated at 65 °C for 1 h with intermittent mixing. After incubation, 3 ml of a 24:1 V/V combination of chloroform and isoamyl alcohol was added and homogenized by gentle inversion for 15 min before centrifugation at 12 × g for 10 min at 4 °C. The clear aqueous phase was transferred to a new centrifuge tube, 2/3 of the volume of ice-cold ethanol was added, and the DNA spool was collected in a 1.5-ml microtube after 30 min at 4 °C. DNA was washed three times with 70% ethanol. After discarding the alcohol, the DNA pellet was air-dried and dissolved in 0.5 ml TE buffer. The quality and quantity of DNA were checked on 0.8% agarose gel stained with ethidium bromide.

Table 1 List of dwarf and tall coconut accessions used for the study and their country of origin

Dwarf accessions		
Name of the accession	Country of origin	No. of palms
Gangabondam Green Dwarf (GBGD)	India	6
Malayan Green Dwarf (MGD)	Malaysia	6
Chowghat Orange Dwarf (COD)	India	6
Malayan Orange Dwarf (MOD)	Malaysia	6
Malayan Yellow Dwarf (MYD)	Malaysia	6
Chowghat Green Dwarf (CGD)	India	6
Cameroon Red Dwarf (CRD)	Cameroon	6
Gudanjali Green Dwarf (GUGD)	India	6
Tall accessions		
Name of the accession	Country of origin	No. of palms
Lakshadweep Ordinary Tall (LCT)	India	6
Philippines Ordinary Tall (PHOT)	The Philippines	6
Fiji Tall (FJT)	Fiji	6
Jamaican San Blas Tall (JSBT)	Jamaica	6
Andaman Ordinary Tall (ADOT)	India	6
San Ramon Tall (SNRT)	The Philippines	6
Laccadive Micro Tall (LMT)	India	6
West Coast Tall (WCT)	India	6

Marker analysis

Ten (10) primer pairs flanking miRNA-SSR and lncRNA-SSR loci were randomly selected for screening. A germplasm panel comprising a single palm of three tall and three dwarf accessions was used for the initial screening of selected SSR primers using PCR. The PCR reactions were conducted in volumes of 20 µl containing 30 ng genomic DNA, 0.2 µM of each primer, 0.1 µl of 10 mM of each dNTPs (MBI Fermentas), 2 µl of 10 X buffer [10 Mm Tris–HCl (pH 8.3), 50 Mm KCl, 1.5 mM $MgCl_2$], and 0.5 µl of *Taq* DNA polymerase [3 units/µl] (Bangalore Genei). The PCR conditions were set to an initial denaturation step at 94 ºC for 2 min, followed by 36 cycles at 94 ºC for 1 min, annealing (different temperature for each primer pair) for 1 min, and 72 ºC for 30 s, terminating with a final extension at 72 ºC for 10 min. Based on the banding pattern, the polymorphic primers were selected and used for the diversity study. PCR products were analyzed using a MultiNA microchip electrophoresis system with DNA 500 kit (MCE-202 MultiNA, Shimadzu). Automated detection of peak size on electropherograms was done by the MultiNA software (MultiNA Control software and Data Analysis software MulltiNA Viewer, Shimadzu).

Data analysis

The GDA (Genetic Data Analysis; Lewis 2001) programme was used to calculate genetic diversity values and create the unweighted pair group method with arithmetic mean (UPGMA) was used to generate a dendrogram using Nei's genetic distance (Nei et al. 1983).

Results

Frequency of miRNA-SSRs in the coconut genome

Based on the results of the previous work (Sabana et al. 2020), 46 miRNA families were selected comprising 110 miRNA members in coconut to develop the miRNA-SSR markers. Using the pre-miRNA prediction tool, 95 precursor sequences of these mature miRNAs were extracted from the CGD genome sequence (Rajesh et al. 2020). The length of the pre-miRNA sequences ranged from 65 to 216 bp and belonged to 29 coconut miRNA families. A total of 95 coconut pre-miRNA sequences were used for mining miRNA-SSRs using the MISA search tool. A total of 88 potential SSRs were identified from 50 pre-miRNA sequences, out of which 23 pre-miRNA genes contained more than one SSR (Table 2). Of these 88 SSRs, 58 (65.90%) comprised simple repeat motifs, whereas 30 (34.10%) were compound motifs. The range of SSR repeats varied from a minimum of mono-nucleotide to a maximum of hexanucleotide. The di-nucleotide repeats were most abundant (62: 70.45%) followed by the tri-nucleotide (8: 9.09%), tetra-nucleotide (6: 6.82%), penta-nucleotide (5: 5.68%), mono-nucleotide (4: 4.55%), and hexa–nucleotide (3: 3.41%) repeats (Table 2).

The frequencies of various SSR motifs are summarized in Fig. 2A. Among the di-nucleotide repeats, AG/CT was the most abundant motif (72.58%), followed by the AT/AT (19.35%) and AC/GT (8.06%). AAG/CTT (37.5%), and ACAT/ATGT (66.67%) were the most abundant repeats in tri-nucleotide and tetra-nucleotide repeats, respectively.

Frequency of lnc-SSRs in the coconut genome

Five hundred nine potential SSRs were identified from 389 lncRNA sequences, out of which 81 lncRNA sequences contained more than one SSR (Table 2). Out of 509 total SSRs, 440 (86.44%) comprised simple repeat motifs, whereas 69 (13.56%) were compound motifs. The range of SSR repeats varied from a minimum of mono-nucleotide to a maximum of hexa-nucleotide. Of these, di-nucleotide repeats were most abundant (318: 62.48%) followed by the tri-nucleotide (136: 26.72%), mono-nucleotide (36: 7.07%), and tetra-nucleotide (16: 3.14%). Similarly, for penta-nucleotide and hexa–nucleotide, one (0.20%) and two (0.39%) repeats were identified, respectively (Table 2).

Similarly, the frequencies of various SSR motifs are summarized in Fig. 2B. Among the di-nucleotide repeats, AG/CT was the most abundant motif (63.21%), followed by the AT/AT (20.13%) and AC/GT (16.35%). AAG/CTT (44.12%) and AAAT/ATTT (31.25%) were the most abundant repeats in trinucleotide and tetra nucleotide repeats, respectively.

Validation of ncRNA-based SSR markers and diversity study

To study the diversity among coconut accessions, 20 SSR primers (10 each from miRNA-SSR and lnc-SSR) were randomly selected and used for genotyping the 96 coconut palms of the 16 accessions (Supplementary Table 1). Out of 20 primers, 10 primers were found to be polymorphic (50%). Of the polymorphic primers, seven were lnc-SSR

Table 2 Details of different SSR motifs of coconut lncRNA-SSR and miRNA-SSR

Item	lnc-SSRs	miRNA-SSRs
Total number of sequences examined	5110	95
Total size of examined sequences (bp)	3,202,509	105,503
Total number of identified SSRs	509	88
Number of SSR containing sequences	389	50
Number of sequences containing more than 1 SSR	81	23
Number of SSRs present in compound formation	69	1830
Repeat type		
Mono-nucleotide	36	4
Di-nucleotide	318	62
Tri-nucleotide	136	8
Tetra-nucleotide	16	6
Penta-nucleotide	1	5
Hexa-nucleotide	2	3

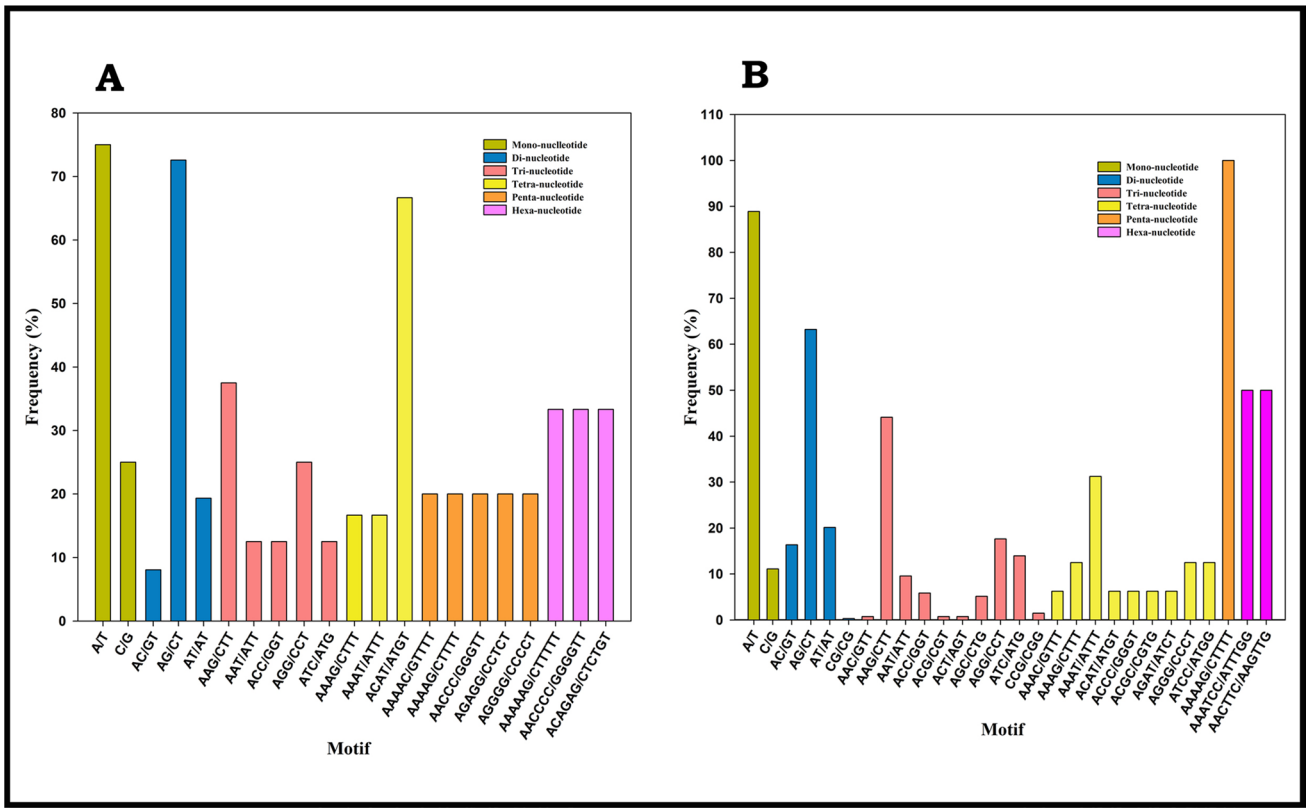

Fig. 2 Frequency of different non-coding RNA-based simple sequence repeat (ncRNA-SSR) markers. (**A**) miRNA-SSR. (**B**) lncRNA-SSR

(lncSSR_53976.1, lncSSR_673, lncSSR_20713.23.1, lncSSR_36364.1, lncSSR_11432.2, lncSSR_37974.1 and lncSSR_38477.1) and three were miRNA-SSR (Cn-miR159c-SSR, Cn-miR393-SSR and Cn-miR394-SSR). The representative digital gel profile of miRNA-SSR and lncRNA-SSR generated by MultiNA Viewer Software is shown in Supplementary Figs. 1 and 2.

The 10 polymorphic SSRs produced 46 alleles ranging from 2 (lncSSR_11432.2) to 9 (Cn-miR393-SSR) alleles per SSR marker, with an average of 4.6 alleles per locus. The effective number of alleles varied from 1.80 to 3.91, with an average of 2.7. Another parameter, Shannon index ranged from 0.64 (lncSSR_11432.2) to 1.41 (lncSSR_53976.1) with a mean value of 1.1. The average observed and expected heterozygosity were 0.5 and 0.62, respectively. The PIC

Table 3 The details of genetic diversity parameters observed in non-coding RNA-based simple sequence repeats (ncRNA-SSRs)

Sl No	SSR	No. of alleles	Effective number of alleles	Shannon's Information Index	Observed heterozygosity (H_o)	Expected heterozygosity (H_e)	PIC value
1	Cn-miR159c-SSR	6	2.91	1.38	0.05	0.66	0.66
2	Cn-miR393-SSR	9	2.77	1.27	0.76	0.64	0.64
3	Cn-miR394-SSR	6	3.46	1.40	1.00	0.71	0.71
4	lncSSR_53976.1	5	3.57	1.41	0.23	0.72	0.72
5	lncSSR_6733.1	3	2.92	1.08	0.51	0.66	0.66
6	lncSSR_36364.1	4	3.91	1.38	0.53	0.75	0.74
7	lncSSR_11432.2	2	1.80	0.64	0.20	0.45	0.44
8	lncSSR_37974.1	3	2.23	0.90	0.80	0.55	0.55
9	lncSSR_38477.1	3	2.06	0.80	0.81	0.52	0.51
10	lncSSR_20713.2	5	2.16	0.91	0.17	0.54	0.54
Mean		4.6	2.78	1.12	0.51	0.62	0.62

values ranged from 0.44 (lncSSR_11432.2) to 0.74 (lncSSR_36364.1), with a mean value of 0.62 (Table 3).

Cluster-based dendrogram and diversity analysis

Genetic variation at 10 ncRNAs loci was assessed in 96 palms of the 16 coconut accessions comprising eight tall and dwarf accessions. The expected heterozygosity (H_e), observed heterozygosity (H_o), Shannon's Information index (I), and fixation index (F_{ST}) in the 16 accessions at 10 active loci are given in Table 4.

The 46 allelic data scored from 10 polymorphic ncRNA-based SSR markers on the 16 coconut accessions were used to generate the dendrogram. The cluster analysis was done using the distance-based cluster analysis (UPGMA), and the 16 accessions were clustered into two main clusters (Fig. 3). All the dwarf and tall accessions formed two different clusters, i.e., I and II, respectively. Within cluster I, Cameroon Red Dwarf formed a separate sub-cluster. In the case of cluster II, San Ramon Tall formed a separate sub-cluster.

Discussion

Molecular markers based on non-coding RNAs are considered informative in the study of complex developmental process and in the adaptation to various physiological stresses. In this study, 597 ncRNA-based SSRs, including 88 miRNA-SSRs and 509 lncRNA-SSRs in coconut, were developed. The frequency of miRNA-SSR is 52.63%, while lncRNA-SSR is 7.61%. Less frequency of lncRNA-SSRs was also reported in capsicum (Jaiswal et al. 2020), whereas in wheat (Bhandawat et al. 2020), it was much higher (44.4%) in comparison to our study. The frequency of miRNA-SSRs is much higher in several plants like rice, wheat, and capsicum (Ganie et al. 2015; Bhandawat et al. 2020; Jaiswal et al. 2020). Di-nucleotide SSRs were more abundant in coconut, followed by tri and tetra-nucleotide SSR motifs, in line with previous reports in capsicum and rice (Ganie et al. 2015; Jaiswal et al. 2020).

We have designed 20 ncRNA-based SSR primers, and 50% of the markers were polymorphic among the 96 coconut palms studied. The level of polymorphism was found to be lower as compared to genomic SSR (74%) but higher compared to EST-SSR (41.7%) in coconut (Riangwon et al. 2020; Preethi et al. 2020). Among these, lncRNA-SSR shows higher polymorphism (70%), which was comparable with the previous report on wheat (Bhandawat et al. 2020), while among miRNA-SSR, 30% (3 markers out of 10) were found to be polymorphic, which was lower. The mean PIC value of 0.62 indicates the ability of ncRNA-SSR markers to detect the high level of polymorphism, and the Nm value (0.81) indicates a high gene flow.

The ncRNA-based SSR markers are being used for genetic diversity studies (Patil et al. 2020; Jaiswal et al. 2020). Among the 16 coconut accessions studied, two distant clusters were obtained in which all eight tall and dwarf accessions are differentiated. Among the tall accessions, San Ramon Tall clustered furthest from all the tall accession and was completely distinct from all the rest. It is worth mentioning that San Ramon Tall, although from the Philippines coconut population, is known for its large fruit size and

Table 4 Expected heterozygosity (H_e), observed heterozygosity (H_o), and fixation index (F_{ST}) across 16 coconut accessions based on the 10 polymorphic ncRNA-SSR markers

Population	Expected heterozygosity (H_e)	Observed heterozygosity (H_o)	Shannon's Information index (I)	Fixation index (F_{ST})
Gangabondam Green Dwarf	0.246970	0.383333	0.3222	−0.642857
Malayan Green Dwarf	0.543939	0.500000	0.8406	0.088146
Chowghat Orange Dwarf	0.384848	0.450000	0.5736	−0.189427
Malayan Orange Dwarf	0.333333	0.533333	0.4525	−0.702128
Malayan Yellow Dwarf	0.327273	0.516667	0.4524	−0.675676
Chowghat Green Dwarf	0.319697	0.516667	0.4368	−0.722222
Cameroon Red Dwarf	0.451515	0.516667	0.6752	−0.161049
Gudanjali Green Dwarf	0.404545	0.550000	0.5650	−0.410256
Lakshadweep Ordinary Tall	0.465051	0.440000	0.6730	0.045509
Philippines Ordinary Tall	0.598485	0.583333	0.9114	0.027778
Fiji Tall	0.583333	0.600000	0.8668	−0.031519
Jamaican San Blas Tall	0.601515	0.500000	0.9032	0.182561
Andaman Ordinary Tall	0.570000	0.583333	0.8281	−0.027719
San Ramon Tall	0.377273	0.483333	0.5608	−0.318182
Laccadive Micro Tall	0.431818	0.400000	0.6877	0.080460
West Coast Tall	0.529675	0.543333	0.7994	−0.028263
Mean	0.448079	0.506250	0.5608	0.146598

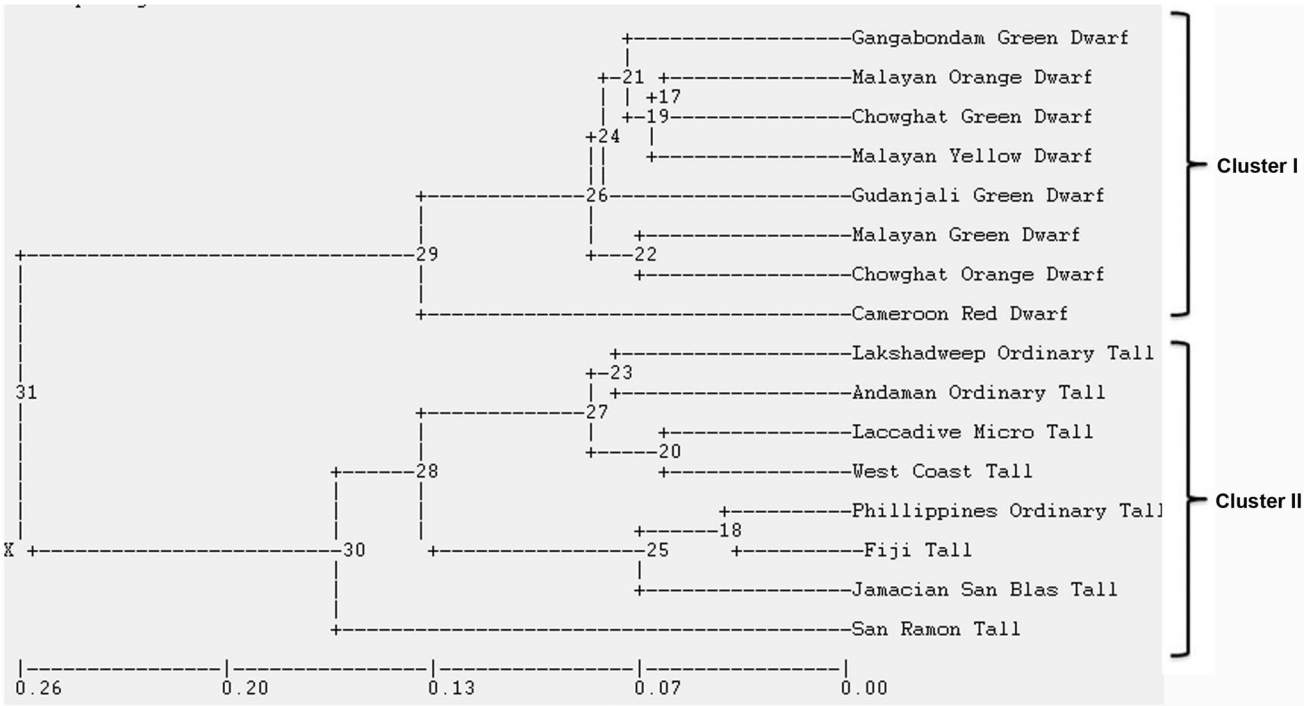

Fig. 3 UPGMA dendrogram showing the genetic relationship among the 16 coconut accessions based on Nei's (1983) genetic distance

morphologically distinct from all other coconut accessions. A selection from San Ramon tall was released as a variety of "Kalpa Shatabdi" from India for commercial cultivation with large fruits, stout stems, and closely arranged leaf scars over the stem. The accession from India, Laccadive Micro Tall, was grouped along with West Coast Tall, whereas Andaman Ordinary Tall was closely related to Lakshadweep Ordinary Tall. Since sailors frequented these two places in the past, it is possible the Lakshadweep coconut population could have developed from the coconut populations of the Andaman and Nicobar Islands. Similarly, the movement of coconut from the West Coast of India, particularly Kerala State, to the Lakshadweep Islands was owing to the close ties between the two regions and human movement, which could account for the grouping of Laccadive Micro Tall with West Coast Tall. Jamaican San Blas Tall, the lone Caribbean coconut accession used in the study, was closely clustered with Fiji Tall and Philippines Ordinary Tall. Among the tall accessions used in the study, improved selections have been released for commercial cultivation from all accessions except Laccadive Micro Tall, a germplasm registered for its unique traits of smaller fruit size, cluster bearing and high kernel oil content (Niral et al. 2014a).

All the dwarf accessions were clustered together except Cameroon Red Dwarf, which formed a single accession sub-cluster. Cameroon Red Dwarf has been registered as a unique dwarf material having higher copra content (Niral et al. 2014b). The clustering pattern revealed that accessions with different colored fruits are clustered together, and the Malayan dwarfs are scattered among the population. Among the indigenous dwarfs, Gangabondam Green Dwarf and Gudanjali Green Dwarf clustered far from the accessions in their sub-cluster, indicating their uniqueness. Gangabondam Green Dwarf has been identified as the best general combiner (Nampoothiri et al. 1999) and has been used to develop several hybrids, whereas Gudanjali Green Dwarf has been known for its early flowering. Although Chowghat Green Dwarf and Gudanjali Green Dwarf were morphologically similar as per observations, the molecular studies indicate that they were different, wherein Chowghat Green Dwarf was closer to Malayan Dwarfs, viz., the Malayan Orange and Yellow Dwarfs. The Malayan Dwarfs clustered with other dwarfs in previous studies (Preethi et al. 2020), as they were the most commonly used dwarf coconut populations worldwide. This suggested that there was a possibility of introgression with many indigenous coconut populations. Selections from all four Malayan dwarfs, Chowghat Orange Dwarf, Chowghat Green Dwarf and Gangabondam Green Dwarf, have been released for commercial cultivation in India, besides their use in hybrid development (Niral et al. 2016).

The SSR markers reported here showed high polymorphism among coconut accessions and the results of the UPGMA analysis revealed that the developed ncRNA-SSR markers are highly suitable for genetic diversity studies across coconut accessions. These ncRNA-SSR markers

can serve as an important resource for functional markers in coconut breeding programs for QTL/gene mapping and association studies for trait dissection.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s10142-022-00911-2.

Author contribution GA, MKR, and AAS designed and executed the project; MKR, AAS, VN, RS, and BAJ collected and prepared the samples and extracted DNA for sequencing; AAS, MKR, KPG, and GA generated and analyzed the data; AAS, GA, and MKR wrote the manuscript with significant intellectual contributions from all authors; VN, RS, BAJ, and KPG revised the manuscript. All authors have read and approved the final manuscript.

Funding This work was supported by funding from the Indian Council of Agricultural Research (ICAR-CPCRI Project No. 1000761030) to MKR and the Central University of Kerala (internal funding) to GA.

Data availability Small RNA-sequence data available in the NCBI SRA database under accession number PRJNA546491.

Declarations

Ethical approval and consent to participate Not applicable.

Consent for publication All authors have provided their consent to publish this article.

Competing interests The authors declare no competing interests.

Human and animal ethics Not applicable.

References

Arunachalam V, Rajesh MK (2017) Coconut genetic diversity, conservation and utilization. In: Ahuja MR, Jain SM (eds) Biodiversity and conservation of woody plants. Springer, Cham, pp 3–36

Bhandawat A, Sharma H, Pundir N, Madhawan A, Roy J (2020) Genome-wide identification and characterization of novel non-coding RNA-derived SSRs in wheat. Mol Biol Rep 47(8):6111–6125

Brake M, Al-Qadumii L, Hamasha H, Migdadi H, Awad A, Haddad N, Sadder MT (2022) Development of SSR markers linked to stress responsive genes along tomato chromosome 3 (*Solanum lycopersicum* L.). BioTech 11(3):34. https://doi.org/10.3390/biotech11030034

Caro RE, Cagayan J, Gardoce RR, Manohar AN, Canama-Salinas AO, Rivera RL, Lantican DV, Galvez HF, Reaño CE (2022) Mining and validation of novel simple sequence repeat (SSR) markers derived from coconut (*Cocos nucifera* L.) genome assembly. J Genet Eng Biotechnol 20(1):1–4. https://doi.org/10.1186/s43141-022-00354-z

Ganie SA, Mondal TK (2015) Genome-wide development of novel miRNA-based microsatellite markers of rice (*Oryza sativa*) for genotyping applications. Mol Breed 35:51

Geethanjali S., Rukmani JA, Rajakumar D, Kadirvel P, Viswanathan PL (2018) Genetic diversity, population structure and association analysis in coconut (*Cocos nucifera* L.) germplasm using SSR markers. Plant Genet Resour 16(2):156–168

Jaiswal V, Rawoof A, Dubey M, Chhapekar SS, Sharma V, Ramchiary N (2020) Development and characterization of non-coding RNA based simple sequence repeat markers in *Capsicum* species. Genomics 112(2):1554–1564

Kanzana G, Zhang Y, Ma T, Liu W, Wu F, Yan Q, Min X, Yan Z, Muvunyi BP, Li J, Zhang Z (2020) Genome-wide development of miRNA-based SSR markers in *Cleistogenes songorica* and analysis of their transferability to Gramineae/non-Gramineae species. J Appl Genet 61:367–377

Lantican DV, Strickler SR, Canama AO, Gardoce RR, Mueller LA, Galvez HF (2019) *De novo* genome sequence assembly of dwarf coconut (*Cocos nucifera* L.'Catigan Green Dwarf') provides insights into genomic variation between coconut types and related palm species. G3: Genes Genomes Genetics 9(8):2377–2393. https://doi.org/10.1534/g3.119.400215

Lebrun P, N'Cho YP, Seguin M, Grivet L, Baudouin L (1998) Genetic diversity in coconut (*Cocos nucifera* L.) revealed by restriction fragment length polymorphism (RFLP) markers. Euphytica 101:103–108

Lewis PO (2001) Genetic Data Analysis: Computer program for the analysis of allelic data Version 1.0 (d16c). http://lewis.eeb.uconn.edu/lewishome/software.html

Manimekalai R, Nagarajan P, Bharathi M, Jayapragasam M, Karun A, Kumaran PM, Parthasarathy VA (2007) RAPD as useful marker for coconut (*Cocos nucifera* L.) DNA analysis. Indian J Hortic 64(1):1–4

Mehta G, Muthusamy SK, Singh GP, Sharma P (2021) Identification and development of novel salt-responsive candidate gene based SSRs (cg-SSRs) and *MIR* gene based SSRs (mir-SSRs) in bread wheat (*Triticum aestivum*). Sci Rep 11(1):1–5

Min X, Zhang Z, Liu Y, Wei X, Liu Z, Wang Y, Liu W (2017) Genome-wide development of microRNA-based SSR markers in *Medicago truncatula* with their transferability analysis and utilization in related legume species. Int J Mol Sci 18(11):2440

Mondal TK, Ganie SA (2014) Identification and characterization of salt responsive miRNA-SSR markers in rice (*Oryza sativa*). Gene 535(2):204–209

Nadeem MA, Nawaz MA, Shahid MQ, Doğan Y, Comertpay G, Yıldız M, Hatipoğlu R, Ahmad F, Alsaleh A, Labhane N, Özkan H (2018) DNA molecular markers in plant breeding: current status and recent advancements in genomic selection and genome editing. Biotechnol Biotechnol Equip 32(2):261–285

Nampoothiri KUK, Kumaran PM, Jerard BA, Ratnambal MJ, Rao EVVB, Parthasarathy VA (1999) Combining ability in coconut (*Cocos nucifera*). CORD 15(2):1–5

Nei M, Tajima F, Tateno Y (1983) Accuracy of estimated phylogenetic trees from molecular data. J Mol Evol 19(2):153–170

Niral V, Ananda KS, Elain Apshara B, Jerard BA, Chowdappa P (2016) Improved varieties of coconut, arecanut and cocoa. ICAR-Central Plantation Crops Research Institute, Kasaragod, India, p 96

Niral V, Jerard BA, Samsudeen K, Nair RV, Jacob PM, Devakumar K, Ratnambal MJ, Kuma-ran PM, Rao EVV, Pillai RV, Thomas George V (2014a) IND 030-Laccadive Micro Tall (IC0430669; INGR13060), a coconut (*Cocos nucifera*) germplasm with cluster bearing heavy bunches of micronuts. Indian J Plant Genet Resour 27(1):73–74

Niral V, Jerard BA, Samsudeen K, Nair RV, Ratnambal MJ, Kumaran PM, Thomas George V (2014b) IND 092 - Cameroon Red Dwarf (IC0598219; INGR13061), a dwarf coconut (*Cocos nucifera*) germplasm with distinct bright orange coloured nuts, higher content of tender nut water and high copra content. Indian J Plant Genet Resour 27(1):74–75

Patil PG, Singh NV, Parashuram S, Bohra A, Mundewadikar DM, Sangnure VR, Babu KD, Sharma J (2020) Genome wide identification, characterization and validation of novel miRNA-based SSR markers in pomegranate (*Punica granatum* L.). Physiol Mol Biol Plants 26(4):683–696. https://doi.org/10.1007/s12298-020-00790-6

Preethi P, Rahman S, Naganeeswaran S, Sabana AA, Gangaraj KP, Jerard BA, Niral V Rajesh MK (2020) Development of EST-SSR markers for genetic diversity analysis in coconut (*Cocos nucifera* L.). Mol Biol Rep 47(12):9385–9397. https://doi.org/10.1007/s11033-020-05981-8

Rajesh MK, Arunachalam V, Nagarajan P, Lebrun P, Samsudeen K, Thamban C (2008) Genetic survey of 10 Indian coconut landraces by simple sequence repeats (SSRs). Sci Hortic 118(4):282–287

Rajesh MK, Chowdappa P, Behera SK, Kasaragod S, Gangaraj KP, Kotimoole CN, Nekrakalaya B, Mohanty V, Sampgod RB, Banerjee G, Prasad TSK (2020) Assembly and annotation of the nuclear and organellar genomes of a dwarf coconut (Chowghat Green Dwarf) possessing enhanced disease resistance. OMICS J Integr Biol 24(12):726–742

Rajesh MK, Fayas TP, Naganeeswaran S, Rachana KE, Bhavyashree U, Sajini KK, Karun A (2016) *De novo* assembly and characterization of global transcriptome of coconut palm (*Cocos nucifera* L.) embryogenic calli using Illumina paired-end sequencing. Protoplasma 253(3):913–928. https://doi.org/10.1007/s00709-015-0856-8

Rajesh MK, Gangurde SS, Pandey MK, Niral V, Sudha R, Jerard BA, Kadke GN, Sabana AA, Muralikrishna KS, Samsudeen K, Karun A, Keshava Prasad TS (2021) Insights on genetic diversity, population structure, and linkage disequilibrium in globally diverse coconut accessions using genotyping-by-sequencing. OMICS J Integr Biol 25(12):796–809

Rajesh MK, Jerard BA, Preethi P, Jacob TR, Fayas TP, Rachana KE, Anitha K (2013) Development of a RAPD-derived SCAR marker associated with tall-type palm trait in coconut. Sci Hortic 150:312–316

Rajesh MK, Karun A, Parthasarathy VA (2019) Coconut biotechnology. In: Nampoothiri KUK, Krishnakumar V, Thampan PK, Nair MA (eds) The coconut palm (Cocos nucifera L.)-research and development perspectives. Springer, Singapore, pp 191–226. https://doi.org/10.1007/978-981-13-2754-4_6

Rajesh MK, Sabana AA, Rachana KE, Rahman S, Jerard BA Karun A (2015) Genetic relationship and diversity among coconut (*Cocos nucifera* L.) accessions revealed through SCoT analysis. 3 Biotech 5(6):999–1006. https://doi.org/10.1007/s13205-015-0304-7

Ramesh SV, Sudha R, Niral V, Rajesh MK (2022) Enhancing genetic gain in coconut: conventional, molecular, and genomics-based breeding approaches. In: Gosal SS, Wani SH (eds) Accelerated plant breeding, vol 4. Springer, Cham, pp 313–357

Riangwong K, Wanchana S, Aesomnuk W, Saensuk C, Nubankoh P, Ruanjaichon V, Kraithong T, Toojinda T, Vanavichit A Arikit S (2020) Mining and validation of novel genotyping-by-sequencing (GBS)-based simple sequence repeats (SSRs) and their application for the estimation of the genetic diversity and population structure of coconuts (*Cocos nucifera* L.) in Thailand. Hort Res 7:156. https://doi.org/10.1038/s41438-020-00374-1

Sabana AA, Rajesh MK, Antony G (2020) Dynamic changes in the expression pattern of miRNAs and associated target genes during coconut somatic embryogenesis. Planta 251(4):1–18

Singh AK, Chaurasia S, Kumar S, Singh R, Kumari J, Yadav MC, Singh N, Gaba S, Jacob SR (2018) Identification, analysis and development of salt responsive candidate gene based SSR markers in wheat. BMC Plant Biol 18(1):1–5

Summanwar A, Basu U, Kav NN, Rahman H (2020) Identification of lncRNAs in response to infection by *Plasmodiophora brassicae* in *Brassica napus* and development of lncRNA-based SSR markers. Genome 64(5):547–566

Tabkhkar N, Rabiei B, Lahiji HS, Chaleshtori MH (2020) Identification of a new set of drought-related miRNA-SSR markers and association analysis under drought stress in rice (*Oryza sativa* L.). Plant Gene 21:100220. https://doi.org/10.1016/j.plgene.2020.100220

Teulat B, Aldam C, Trehin R, Lebrun P, Barker JH, Arnold GM, Karp A, Baudouin L, Rognon F (2000) An analysis of genetic diversity in coconut (*Cocos nucifera*) populations from across the geographic range using sequence-tagged microsatellites (SSRs) and AFLPs. Theor Appl Genet 100(5):764–771

Tyagi S, Kumar A, Gautam T, Pandey R, Rustgi S, Mir RR (2021) Development and use of miRNA-derived SSR markers for the study of genetic diversity, population structure, and characterization of genotypes for breeding heat tolerant wheat varieties. PLoS ONE 16(2):e0231063

Upadhyay A, Jayadev K, Manimekalai R, Parthasarathy VA (2004) Genetic relationship and diversity in Indian coconut accessions based on RAPD markers. Sci Hortic 99(3–4):353–362

Wang X, Gui S, Pan L, Hu J, Ding Y (2016) Development and characterization of polymorphic microRNA-based microsatellite markers in *Nelumbo nucifera* (Nelumbonaceae). Appl Plant Sci 4(1):1500091

Xiao Y, Xu P, Fan H, Baudouin L, Xia W, Bocs S, Xu J, Li Q, Guo A, Zhou L, Li J (2017) The genome draft of coconut (*Cocos nucifera*). Giga Sci 6(11):gix095. https://doi.org/10.1093/gigascience/gix095

Yang Y, Bocs S, Fan H, Armero A, Baudouin L, Xu P, Xu J, This D, Hamelin C, Iqbal A, Qadri R (2021) Coconut genome assembly enables evolutionary analysis of palms and highlights signaling pathways involved in salt tolerance. Commun Biol 4(1):1–4