

# Genetic diversity, population structure and association analysis in coconut (*Cocos nucifera* L.) germplasm using SSR markers

S. Geethanjali<sup>1\*</sup>†, J. Anitha Rukmani<sup>1,2</sup>, D. Rajakumar<sup>1</sup>, P. Kadirvel<sup>3</sup> and P.L. Viswanathan<sup>2</sup>

<sup>1</sup>Coconut Research Station, Tamil Nadu Agricultural University, Aliyarnagar-642101, Coimbatore, India,

<sup>2</sup>Department of Oilseeds, Tamil Nadu Agricultural University, Coimbatore-641003, India and <sup>3</sup>ICAR-Indian Institute of Oilseeds Research, Rajendranagar, Hyderabad-500030, India

Received 19 July 2016; Accepted 24 February 2017 – First published online 10 April 2017

## Abstract

A world-wide coconut germplasm collection (79 genotypes) was analyzed for genetic diversity and population structure based on 48 simple sequence repeat (SSR) loci. The genotypes displayed moderately high amount of genetic diversity, which was strongly structured according to geographical origins. Number of SSR alleles ranged from 2 to 7 with an average of 4.1 per locus. Gene diversity (expected heterozygosity) estimates ranged from 0.162 to 0.811 with a mean of 0.573. Polymorphism information content values ranged from 0.149 to 0.785 with an average of 0.522. Hierarchical clustering analysis grouped the genotypes into two major clusters with two sub-groups in each, which corresponded with the geographic origins. The first cluster comprised of ‘Tall’ genotypes originated from Indo-Atlantic and South Asia regions. The second cluster comprised mostly of ‘Dwarf’ genotypes and some Tall genotypes which originated from Indo-Pacific and South-East Asia regions. Model-based clustering by STRUCTURE analysis also supported the presence of clear genetic structuring in the collection with two major populations ( $K=2$ ) and four sub-populations ( $K=4$ ). The proportion of SSR locus-pairs in linkage disequilibrium was low (2.4%). Association analysis in a subset of 44 genotypes detected a single SSR locus, CnCir73 (chromosome 1) putatively associated with fruit yield component traits, which corresponded with a previously mapped quantitative trait locus in coconut.

**Keywords:** hierarchical clustering, linkage disequilibrium, microsatellite markers, plantation crop

## Introduction

Coconut (*Cocos nucifera* L.) (Arecaceae) is a globally important tree-borne oil crop. It is grown in more than 80 countries, while Indonesia, the Philippines and India are the major producers contributing about 75% of the total production (FAOSTAT, 2012). Understanding of genetic

diversity in germplasm collections would be helpful for conservation, maintenance and exploitation of this valuable species. Coconut germplasm is primarily classified into ‘Tall’ (*Typica*) and ‘Dwarf’ (*Nana*) types. The Talls are outcrossing and the Dwarfs are mostly autogamous in nature (Harries, 1978). Traditionally, coconut genotypes are characterized using morphological traits (Balakrishnan and Nair, 1979; Panda, 1982; Sugimura et al., 1997; Zizumbo-Villarreal and Piñero, 1998), which have been difficult due to scarcity of phenotypic markers, long juvenile phase and expensive field trials. Molecular markers help to overcome these limitations as they are abundant, highly

\*Corresponding author. E-mail: [geethanjalitnau@yahoo.com](mailto:geethanjalitnau@yahoo.com)

†Current address: Centre for Plant Breeding and Genetics, Tamil Nadu Agricultural University, Coimbatore-641003, India.

polymorphic, independent of environments and the characterization can be performed rapidly (Perera et al., 1998). Several kinds of molecular markers have been used to characterize genetic diversity in coconut namely randomly amplified polymorphic DNA (RAPD) (Ashburner et al., 1997; Daher et al., 2002), RFLPs (restriction fragment length polymorphisms) (Lebrun et al., 1998), AFLPs (amplified fragment length polymorphisms) (Perera et al., 1998; Teulat et al., 2000), simple sequence repeats (SSRs) (Rivera et al., 1999; Perera et al., 2000, 2001; Teulat et al., 2000; Baudouin and Lebrun, 2002; Meerow et al., 2003; Rajesh et al., 2008a, b; Martinez et al., 2010; Ribeiro et al., 2010; Liu et al., 2011; Loiola et al., 2016), inter-SSRs (Manimekalai and Nagarajan, 2006) and SNPs (single-nucleotide polymorphisms) (Mauro-Herrera et al., 2007). Among all, SSRs are considered ideal markers because of locus specificity, multi-allelic and codominant nature (Tautz, 1989), which are critical properties for assessing genetic diversity in outcrossing species.

Germplasm collections are directly exploited for association mapping of traits in perennial species (Font i Forcada et al., 2015). Traditional linkage mapping has limitations in perennial species because it requires pedigree-based mapping population such as recombinant inbred lines produced from cross between two diverse inbreds, which is extremely difficult to develop (Herrán et al., 2000; Baudouin et al., 2006; Riedel et al., 2009). A germplasm collection which displays excellent trait variability, genetically unstructured and free of linkage disequilibrium (LD) among unlinked loci is considered ideal for association analysis. Studies on association mapping of traits in coconut germplasm are rare. To date, a few preliminary association studies have been reported in coconut, which include association of RAPD markers with resistance to lethal yellowing disease (Cardena et al., 2003), resistance to eriophyid mite (Shalini et al., 2007) and gene-based SSR markers with tree height (Xia et al., 2014). In this study, our objectives were to analyse genetic diversity, population structure and LD in a world-wide collection of coconut genotypes (79), maintained at a regional coconut breeding station in Tamil Nadu state (India), using SSR markers and subsequently to perform an association analysis with fruit yield components in a subset of 44 genotypes.

## Materials and methods

### Plant material

In this study, a total of 79 coconut genotypes (34 exotic and 45 indigenous) were used (Table 1). Details of source/origin of these genotypes are provided in Table S1. The exotic genotypes represented 15 countries including Africa (5), Fiji Islands (2), Guam Islands (2), Indonesia (2), Jamaican

Islands (1), Malaysia (4), Micronesia (1), Panama (1), Papua New Guinea (1), the Philippines (8), Thailand (2), Trinidad (1), Solomon Islands (1), Sri Lanka (2) and Vietnam (1). The indigenous genotypes represented different parts of India: Andaman and Nicobar Islands (5), Andhra Pradesh (1), Goa (3), Karnataka (2), Kerala (4), Lakshadweep Islands (2), Tamil Nadu (27) and West Bengal (1). Fifty genotypes were maintained at Coconut Research Station (CRS), Tamil Nadu Agricultural University (TNAU), Aliyarnagar (10.4881514°N, 76.9657343°E), Tamil Nadu State, India; 13 at CRS, TNAU, Veppankulam (10.44895°N, 79.34915°E), Tamil Nadu State, India and 12 were from farmers' fields located at three districts namely Nagercoil (8.1700°N, 77.4300°E), Theni (10.0090°N, 77.4700°E) and Coimbatore (11.0183°N, 76.9725°E) of Tamil Nadu State, India and four from Coconut Nursery, TNAU, Coimbatore, India. The coconut genotypes were established at TNAU with the support of Indian Council of Agricultural Research (ICAR), New Delhi through its All India Coordinated Research Programme on Palms (AICRP-Palms).

### Measurements on fruit traits

A subset of 44 genotypes maintained at CRS, TNAU, Aliyarnagar, Tamil Nadu State, India was of uniform age (22 years old); hence, it was considered for phenotypic measurements. Evaluation of trees for fruit component traits was carried out for a period of 3 years (2011–2013) consecutively. For each genotype, observations on fruit traits were recorded (12 nuts/palm/year) as per the standard descriptors (Ratnambal et al., 1995), which are briefly described below.

- i. Fruit length – length between the two polar zones in cm.
- ii. Fruit breadth – measured at the equatorial zone of the fruit in cm.
- iii. Fruit weight – weight of the whole fruit (10-month old) in g.
- iv. Husk thickness – measured from the pericarp of the fruit to the outer shell of the nut in cm.
- v. Nut weight – weight of the dehusked nut in g.
- vi. Kernel weight – weight of the kernel after separation from the shell in g.
- vii. Kernel thickness – measured in split nuts in cm.
- viii. Shell thickness – measured in split nuts in cm.
- ix. Copra content – weight of the copra in g.
- x. Nut yield – number of matured nuts per palm per year.

### DNA extraction

DNA isolation was carried out using a high salt concentration method described by Angeles et al. (2005) with some modifications in the sample collection and initial extraction conditions. Two to three fresh whole leaflets were collected

**Table 1.** Details of 79 coconut genotypes used in this study

Region	Countries/provinces	Genotypes <sup>a</sup>	
		Tall	Dwarf
Indo-Atlantic	Africa	<b>KNT, ZAT, SCT, GOT, NIT</b>	–
	Jamaica	<b>JMT</b>	–
	Panama	<b>PNT</b>	–
	Trinidad and Tobago	<b>STVT</b>	–
Oceania	Fiji Islands	<b>FJT, FJRT</b>	–
	New Guinea	<b>NGAT</b>	–
	Micronesia	<b>FMST</b>	–
	Solomon Islands	<b>BSIT</b>	–
South-East Asia	Guam Islands	<b>GUT-1, GUT-2</b>	–
	Indonesia	<b>BRT, JVT</b>	–
	Thailand	<b>TALT, SIT</b>	–
	The Philippines	<b>PLT-1, PLT-2, PHOT-1, PHOT-2, PHOT-3, PHOT-4, SNRT-1, SNRT-2</b>	–
South Asia	Vietnam	<b>CCNT</b>	–
	Malaysia	<b>SSGT</b>	<b>MGD, MOD, MYD</b>
	Sri Lanka	<b>SLT, GTBT</b>	–
	India		
	Andaman Islands	<b>AGT, ADOT, ADRT</b>	AOD
	Car Nicobar Islands	<b>NICT</b>	–
	Lakshadweep Islands	<b>LMT, LCT</b>	–
	Andhra Pradesh	–	<b>GBGD</b>
	West Bengal	<b>HZT</b>	–
	Goa	<b>NDRT-1, NDRT-2, CALT</b>	–
	Karnataka	<b>TPT</b>	<b>KTOD</b>
	Kerala	<b>KPDT, SPIT</b>	<b>CGD, COD</b>
	Tamil Nadu	<b>APT, AYRT-1, AYRT-2, ECT-1, ECT-2, EMT, RJPT, WCT</b>	KGD, KYD
	Local ecotypes from farmer fields in Tamil Nadu		
	Thanjavur	<b>ALT, THMT, VPMT, MPMT</b>	–
	Coimbatore	<b>THT-F, POLT-F-1, POLT-F-2</b>	–
	Nagapattinam	<b>VELT</b>	–
Theni	<b>CMT-F, GPT-F</b>	–	
Kanyakumari	<b>EMT-F-1, EMT-F-2, EMT-F-3, EMT-F-4, EMT-F-5, VKT-F-1, VKT-F-2</b>	–	

<sup>a</sup>Genotypes were coded and actual names are provided in supplementary information (Table S1); the genotypes in bold were used for association analysis.

from the second topmost frond without any damage. The samples from different locations were packed in ziplock covers and transported to laboratory under room temperature condition. The samples remained green and fresh for about 3 d. A day before DNA extraction, pestle and mortars were placed in deep freezer (–80°C) for pre-chilling. About 10 ml of the DNA extraction buffer was added to 50 ml falcon tubes (sterilized) and pre-heated to 65°C in water bath. The fresh leaflets were cut into small bits of about 5 cm

length and were quickly dropped into the pre-chilled pestle and mortar (–80°C) containing liquid nitrogen. Care was taken to avoid browning of leaf samples. Immediately, 500 mg of PVP (polyvinylpyrrolidone) was added and the leaf samples were ground to fine powder. The powdered samples were quickly transferred (before thawing) to the falcon tubes containing 10 ml of pre-heated DNA extraction buffer. Then, 1 ml of 20% SDS (sodium dodecyl sulphate) and 50 µl of β-mercaptoethanol were added. The

contents were mixed thoroughly, incubated at 65°C for 1 h at water bath and centrifuged at 15,000 rpm for 15 min at 4°C. The supernatants were carefully transferred to fresh falcon tubes. For every 1 ml of supernatant, 250 µl of 3 M sodium acetate and 1 ml of ice cold absolute isopropanol were added. The resulting solution was mixed by gentle inversion followed by incubation for 20 min at –20°C for DNA precipitation to happen. The DNA was hooked out and washed with 70% ethanol twice and suspended in 1 ml of TE buffer. The DNA was further purified by adding about 100 µl of 3 M sodium acetate and 2 ml of pre-chilled absolute alcohol, mixing gently and incubating at 4°C for 30 min. Then, the DNA was hooked out or pelleted, washed with 70% ethanol, air dried for 2 h and dissolved in TE buffer before storing at –20°C until further use. The quality and quantity of DNA were checked by loading 3 µl of DNA sample on 1% agarose gel along with the standard  $\lambda$  DNA marker with ethidium bromide staining. Based on the results of quantification, the stock DNA of all samples were diluted using sterile distilled water to have working samples with approximately 10 ng/µl concentration for SSR analysis.

### SSR analysis

A set of 48 SSR primer pairs was selected for genotyping work from the Tropgene database (<http://www.cogentnetwork.org/tropgene-database>), which was developed by Centre International de Recherches en Agronomie pour le Développement (CIRAD), Montpellier, France. These primers with known linkage groups (LG) and map positions were selected in order to have a random coverage of markers throughout 16 chromosomes of coconut. Number of markers on each LG was as follows: LG1-5, LG2-5, LG3-2, LG4-5, LG5-1, LG7-1, LG8-4, LG9-3, LG10-4, LG11-4, LG12-2, LG13-7, LG14-1, LG15-3 and LG16-1. The details of SSR loci, genetic map positions and the sequences of primer-pairs are provided in Table S2.

The PCR was carried out in 20 µl volume containing 4 µl of genomic DNA (10 ng/µl), 2.0 µl of 10X *Taq* buffer containing 1.5 mM MgCl<sub>2</sub> (20 mM stock), 2.0 µl of dNTP (2.5 mM of each dNTP), 1.0 µl each of forward and reverse primer (100 pM/µl stock), 0.5 µl of *Taq* DNA polymerase (3U/µl), and 9.5 µl of sterile water. PCR cycles consisted of an initial denaturation step at 94°C for 5 min, followed by 35 cycles at 94°C for 30 s for denaturation, 1 min at 51°C for primer annealing, and 1 min at 72°C for extension. This was followed by a final extension step at 72°C for 5 min. About 3 µl of PCR products were run on 6% polyacrylamide gel electrophoresis (PAGE) at 150 volts for 3 h and resolved by silver staining procedure described by Panaud et al. (1996). The SSR alleles were scored as allele 1, 2, 3, etc. based on size (bp) (approximate).

### Statistical analysis

#### Basic statistics of trait variation

Mean, range and standard deviation (SD) for each trait were obtained using the 'Data Analysis Tools' available in the standard Microsoft Excel program.

#### Genetic diversity estimates

Number of alleles ( $N_A$ ), frequency of major alleles ( $M_{AF}$ ), gene diversity (expected heterozygosity,  $He$ ) and polymorphism information content (PIC) were obtained using the software program PowerMarker version 3.25 (Liu and Muse, 2005).

#### Cluster analysis

Hierarchical clustering based on Ward's method (Ward, 1963) was used to group genotypes as implemented in DARwin software version 5 (Dissimilarity Analysis and Representation for windows) V.5.0.158 (Perrier and Jacquemoud-Collet, 2006) (available at <http://darwin.cirad.fr/>).

#### Population structure

STRUCTURE 2.3.2 program (Pritchard et al., 2000) was used to detect the number of populations within a collection of 79 coconut genotypes. The membership of each genotype was tested from  $K=1$  to  $K=10$  ( $K$ =putative number of populations) with the admixture model, along with a burn-in period of 100,000 and a run length of 100,000. Subsequently, the  $\Delta K$  measure (Evanno et al., 2005) was used to estimate the optimum number of populations using the online version of Structure harvester ([http://tayloro.biologyucla.edu/Struct\\_harvest](http://tayloro.biologyucla.edu/Struct_harvest)) developed by Earl and von Holdt (2012). A value of  $K$  was selected when the estimate of  $\Delta K$  peaked in the range of 1–10 and inferred ancestry estimates of individuals (Q-matrix) (membership coefficients) were derived for the selected population (Pritchard et al., 2000). The genotypes with membership coefficient  $\geq 0.75$  were assigned to the respective population and  $< 0.75$  were assigned to the admixture group.

#### Estimation of LD between SSR loci

The statistic, squared allele frequency correlation ( $r^2$ ) was used to determine LD between SSR loci (pairwise) as implemented in the software program PowerMarker version 3.25 (Liu and Muse, 2005) with 1000 permutations. The LD between loci was considered significant when  $r^2$  values were  $> 0.1$  at  $P < 0.0001$ . The  $r^2$  values range from 0 to 1.

When  $r^2 = 0$ , two loci are presumed to be in complete linkage equilibrium and when  $r^2 = 1$ , the loci are presumed to be in complete LD. Arbitrarily, LD between two SSR loci is considered significant when  $r^2$  value exceeds 0.1 (Abdurakhmonov and Abdulkarimov, 2008).

### Association analysis

A subset of 44 coconut genotypes (collection of CRS, Aliyarnagar) was considered for association analysis because the trees were of uniform age. Association analysis was carried out based on general linear model (GLM) as implemented in TASSEL (Bradbury et al., 2007). The population membership estimates (Q-matrix) obtained from the Structure software was combined as covariate in the association analysis. This is helpful specially to account for population structure effect on the association analysis. Correction for multiple testing was applied to *P*-values obtained from GLM using the positive false discovery rate (FDR) method as implemented in software program QVALUE (Storey and Tibshirani, 2003). The *q*-value is a measure of significance in terms of FDR similar to the *P*-value that relates to the false-positive rate. Marker trait associations having  $q \leq 0.1$  (arbitrary) were declared significant (Iquiria et al., 2015). The *P* and *q* values determine whether a locus is associated with the marker and the marker  $R^2$  evaluates the magnitude of the locus effect (Agrama et al., 2007).

## Results

### Trait variation

Fruit component traits of coconut namely fruit length, breadth, weight, husk thickness, nut weight, kernel weight,

kernel thickness, shell thickness, copra content and nut yield displayed quantitative variation in the collection of 44 coconut genotypes. Mean, range and SD of traits in the coconut genotype panel from the evaluation trial for three consecutive years (2011–2013) are presented in Table 2.

### Genetic diversity

All 48 SSR loci were polymorphic in the panel of 79 coconut genotypes. A total of 197 alleles was detected with an average of 4.1 alleles per locus. The  $N_A$  at each locus ranged from 2 (CnCirE1, CnCirD1, CnCirI4, CnCirG4, CnCir126) to 7 (CnCirC5, CnCir87, CnCirA4). Allele size (bp) (approximate) information for each SSR locus are provided in Table S2. The mean gene diversity was 0.573 and it ranged from 0.162 (CnCirE1) to 0.811 (CnCirC5). The average PIC value was 0.522 and it ranged from 0.149 (CnCirE1) to 0.785 (CnCirC5). Details of the genetic diversity measures are presented in Table 3.

### Genotypic clusters

Cluster analysis revealed two major clusters (C1 and C2) with four distinct sub-clusters (Fig. 1). The first group (C1) comprised of the Tall genotypes distinctly. Within this group, two sub-clusters (C1.1 and C1.2) could be distinguished. The first sub-cluster (C1.1) comprised of 26 Tall genotypes, which originated from the Indo-Atlantic Region (Laccadive Islands, Andaman and Nicobar Islands, Seychelles, Zanzibar, Tanzania), Caribbean Islands (San Blas and St. Vincent) and the South Asian countries namely Indonesia, Malaysia and Vietnam. The second sub-cluster (C1.2) consisted of 20 Tall genotypes,

**Table 2.** Variation for fruit component traits in a panel of 44 coconut genotypes based on evaluation data over three consecutive years

Trait	2011			2012			2013		
	Mean	Range	SD	Mean	Range	SD	Mean	Range	SD
Fruit length (cm)	26.1	20.3–31.0	2.70	25.2	21.0–30.3	2.50	26.5	20.8–34.0	3.8
Fruit breadth (cm)	17.4	13.8–22.0	1.90	17.2	12.5–23.0	2.20	17.5	12.3–25.3	3.0
Fruit weight (g)	1598.1	966.7–3081.3	481.7	1671.0	1022.5–2665.0	433.5	1572.0	938.3–2681.5	483.7
Husk thickness (cm)	2.68	1.27–4.00	0.69	2.36	0.98–3.75	0.58	2.70	1.40–4.08	0.73
Nut weight (g)	657.6	312.5–1312.5	215.5	736.8	422.5–1427.5	240.3	687.5	372.0–1320.5	224.1
Kernel thickness (cm)	1.24	0.98–1.48	0.12	1.31	1.0–1.55	0.14	1.28	1.00–1.50	0.12
Kernel weight (g)	280.1	108.8–500.0	82.7	330.9	225.0–570.0	86.6	303.9	169.3–533.8	80.4
Shell thickness (cm)	0.24	0.20–0.35	0.05	0.26	0.20–0.50	0.06	0.26	0.20–0.40	0.06
Copra content (g)	159.9	93.2–252.5	36.9	174.2	87.5–272.5	46.9	157.2	85.0–272.5	45.0
Nut yield (g)	111.2	35.0–231.2	42.7	86.7	38.8–141.0	24.5	96.9	19.0–189.0	44.6

**Table 3.** Measures of genetic diversity based on 48 SSR loci in a collection of 79 coconut genotypes

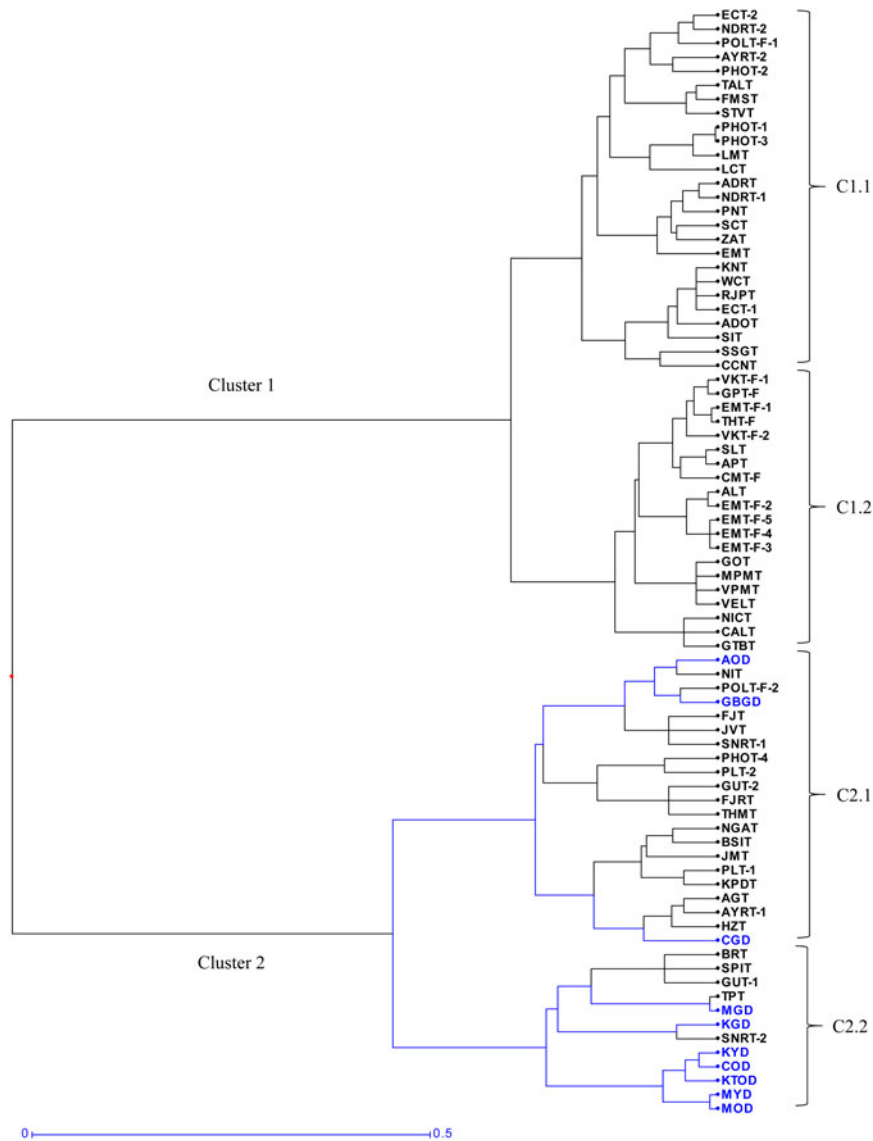
SSR locus	$N_A$	$M_{AF}$	$He$	PIC	SSR locus	$N_A$	$M_{AF}$	$He$	PIC
CnCir202	6	0.633	0.570	0.546	CnCir179	4	0.823	0.310	0.292
CnCirE4	3	0.481	0.579	0.489	CnCir47	5	0.342	0.746	0.704
CnCir226	3	0.544	0.563	0.478	CnCirE7	3	0.798	0.344	0.318
CnCir119	5	0.519	0.651	0.606	CnCir87	7	0.380	0.730	0.685
CnCir73	5	0.633	0.564	0.534	CnCirK8	4	0.506	0.630	0.566
CnCir206	4	0.468	0.695	0.654	CnCirG4	2	0.532	0.520	0.409
CnCir147	2	0.848	0.269	0.253	CnCir93	4	0.544	0.572	0.492
CnCir215	4	0.582	0.548	0.471	CnCirA9	6	0.468	0.708	0.674
CnCir51	3	0.785	0.352	0.312	CnCirI4	2	0.532	0.540	0.439
CnCir151	4	0.468	0.650	0.588	CnCir126	2	0.760	0.376	0.322
CnCir1	3	0.658	0.503	0.446	CnCirC11	6	0.329	0.780	0.748
CnCirB12	4	0.684	0.477	0.424	CnCir74	2	0.658	0.473	0.392
CnCirC9	3	0.481	0.576	0.483	Mean	4.1	0.553	0.573	0.522
CnCir48	4	0.342	0.756	0.716	Range	2–7	0.329–0.911	0.162–0.811	0.149–0.785
CnCirE1	2	0.911	0.162	0.149					
CnCirC5	7	0.266	0.811	0.785					
CnCirD1	2	0.747	0.390	0.331					
CnCir167	3	0.747	0.399	0.350					
CnCir24	3	0.633	0.546	0.502					
CnCir146	3	0.494	0.547	0.445					
CnCir56	6	0.570	0.608	0.563					
CnCirD8	3	0.506	0.615	0.543					
CnCir121	3	0.722	0.442	0.402					
CnCir89	3	0.443	0.650	0.582					
CnCir192	6	0.354	0.763	0.727					
CnCir86	6	0.266	0.771	0.733					
CnCirB11	6	0.734	0.446	0.431					
CnCir2	6	0.722	0.457	0.434					
CnCirJ10	4	0.456	0.661	0.601					
CnCir109	6	0.430	0.697	0.649					
CnCir123	6	0.279	0.786	0.753					
CnCir241	3	0.557	0.569	0.490					
CnCirE11	4	0.519	0.651	0.604					
CnCirB4	4	0.570	0.596	0.542					
CnCirK1	4	0.354	0.750	0.708					
CnCirA4	7	0.443	0.714	0.674					

$N_A$ , number of alleles per SSR locus;  $M_{AF}$ , major allele frequency;  $He$ , expected heterozygosity; PIC, polymorphism information content.

which were mostly of local ecotypes of Tamil Nadu (India), a genotype from the Car Nicobar Islands (India) and all genotypes of Sri Lankan origin used in this study.

The second group (C2) included all Dwarfs and Talls from the Indian Ocean, Oceania and South Pacific Islands, which were further grouped into two distinct sub-clusters (C2.1 and C2.2). The sub-cluster (C2.1) included 21 genotypes (18 Talls and 3 Dwarfs). The sub-cluster (C2.2)

comprised of 12 genotypes (7 Dwarfs and 5 Talls). Most of the orange and yellow coloured Dwarfs (Malayan Yellow Dwarf, Malayan Orange Dwarf, Chowghat Orange Dwarf, Kulasekaram Yellow Dwarf, Kenthali Orange Dwarf) except Andaman Orange Dwarf and the green Dwarfs (Ganga bondam, Chowghat Green Dwarf) grouped distinctly in a single sub-cluster (C2.2), irrespective of their geographic origin. The Talls included mostly the genotypes



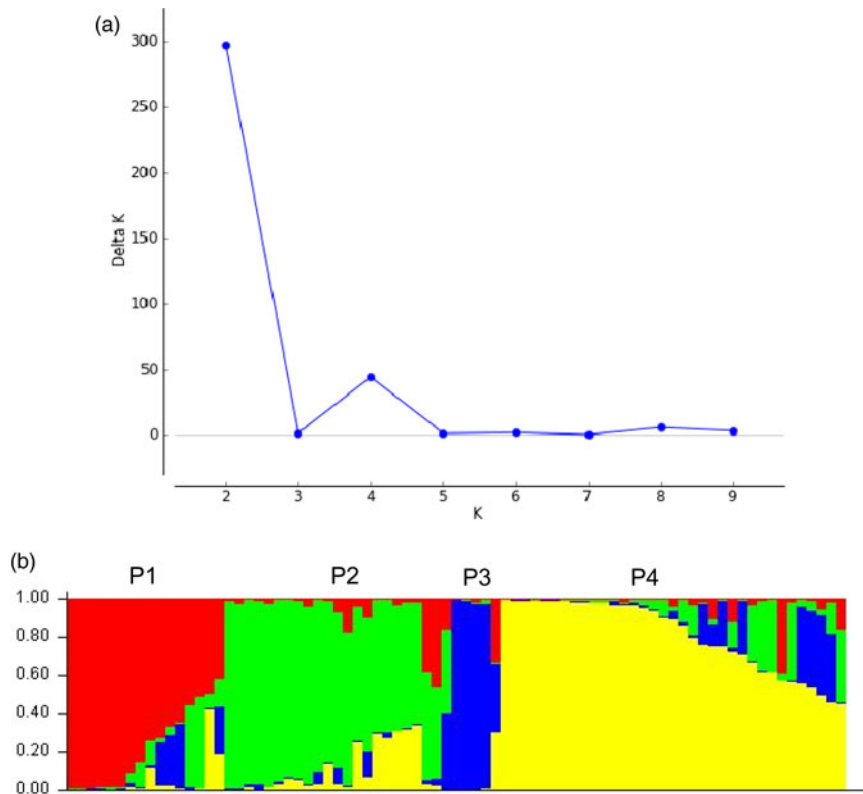
**Fig. 1.** Dendrogram showing genetic relationships of 79 coconut genotypes using 48 SSR loci based on Ward's method of Hierarchical clustering. Cluster 1 included genotypes from Indo-Atlantic to South Asia regions and cluster 2 included genotypes from Indo-Pacific to South-East Asia regions. C1.1, C1.2, C2.1 and C2.2 indicate sub-clusters. Dwarf types are indicated in blue colour.

of Indian origin, the Philippines, Oceania (Melanesia, Micronesia and Polynesia) and South Pacific Islands (Fiji, Guam, New Guinea, and British Solomon Islands), which are towards the Australian continent.

### Population structure

While performing STRUCTURE analysis, 50 data sets were obtained by setting the number of possible clusters ( $K$ ) from 1 to 10 with five replications each. The mean posterior probability [LnP(D)] value for each given  $K$  increased with the increase of  $K$ , but no abrupt change in LnP(D) values

was observed, hence the probable  $K$  value could not be inferred. However, the  $\Delta K$  analysis (Evanno et al., 2005), showed a sharp peak of  $\Delta K$  at  $K=2$  (Fig. 2), suggesting two major populations. A secondary peak of  $\Delta K$  was also observed at  $K=4$  (Fig. 2) indicating further sub-structuring, which were named as sub-populations P1, P2, P3 and P4. The genotypes were assigned to specific population based on the threshold value of membership coefficients ( $\geq 0.75$ ). The values of membership coefficient for each genotype are presented in Table S1. Out of 79 genotypes, 47 could be assigned to either P1, P2, P3 or P4 unambiguously and the remaining 32 genotypes were considered admixture types. P1 included eight genotypes (COD, KTOD, KGD,



**Fig. 2.** Prediction of sub-populations ( $K = 4$ ) in a collection of 79 coconut genotypes using Evanno method in the STRUCTURE analysis (a) and STRUCTURE plot depicting sub-populations (b). P1, P2, P3 and P4 indicate sub-populations.

KYD, MGD, MOD, MYD and TPT) from South Asia, which were mostly Dwarf with the exception of one. P2 included 13 Tall genotypes (AGT, ADOT, AYRT-1, GUT-2, HZT, JMT, KPDT, KNT, NDRT-2, PHOT-4, PNT, TALT and ZAT), which represented Indo-Atlantic, South-East Asia and South Asia. P3 included four Tall genotypes (FJT, GUT-1, PLT-2 and PHOT-2), which represented Oceania and South-East Asia. P4 included 22 genotypes (ALT, APT, CALT, SLT, ECT-1, EMT-F-2, EMT-F-3, EMT-F-4, EMT-F-5, GPT-F, GTBT, LMT, NICT, PHOT-3, POLT-F-1, SNRT-2, SIT, THT-F, VKT-F-1, VKT-F-2, VELT and VPMT), which were mostly from India and the Philippines. The STRUCTURE plot for  $K = 4$  is presented in Fig. 2.

### LD between SSR loci

Overall, a total of 1128 SSR locus-pairs were analyzed for LD. The  $r^2$  ranged from 0.0001 to 0.2283 with a mean of 0.0247. Out of 1128 marker pairs, 27 were in LD across LG with a mean  $r^2$  of 0.1480, which indicated that the LD among SSR loci tested in the population was low (2.4%). Two closely located locus-pairs, on chromosome 1 (CnCirE4-CnCir226) and 2 (CnCir147-CnCir215) showed significant LD with a mean  $r^2$  of 0.1401 as expected.

### Association of SSR marker with fruit component traits

The SSR locus CnCir73 (LG1) showed putative association with fruit breadth, kernel weight, nut weight and copra content at FDR ( $q$ -value)  $\leq 0.1$  with the phenotypic contribution ( $R^2$ ) ranging from 34 to 43% (Table 4).

**Table 4.** SSR marker putatively associated with fruit component traits in coconut germplasm through association analysis

SSR locus	Traits	$R^2$ (%)	$P$ -value	$q$ -value (FDR)
CnCir73 (LG1)	Fruit breadth	0.43	0.00016	0.0554
	Kernel weight	0.39	0.00038	0.0665
	Nut weight	0.35	0.00096	0.1059
	Copra content	0.34	0.00180	0.1059

LG, linkage group;  $R^2$ , phenotypic variance explained by the SSR locus; FDR, false discovery rate of  $\leq 0.1$  was considered for putative association of marker locus with the traits.



## Discussion

In this study, SSR markers revealed that a world-wide coconut germplasm collection (79 genotypes) maintained at Tamil Nadu state (India) preserved moderately high level of genetic diversity and was strongly structured according to geographical origins. Number of SSR alleles ranged from 2 to 7 with an average of 4.1 alleles per locus, which corresponded with previous studies. Perera et al. (2001) reported 3–10 alleles with an average of seven alleles per locus in 33 coconut genotypes based on eight SSR loci. Rajesh et al. (2008a) reported 3–17 alleles with an average of 6.4 alleles per locus in 102 coconut palms based on 14 SSR loci. Dasanayaka et al. (2009) identified a total of 79 alleles (range 3–10, average 4.9 alleles per locus) at 16 SSR loci in 43 coconut genotypes. Ribeiro et al. (2010) reported an average of 5.2 alleles per locus in 10 coconut populations based on 13 SSR loci. Overall, SSR analysis in coconut has revealed a minimum of 2 to a maximum of 17 alleles per locus (Teulat et al., 2000; Rajesh et al., 2008b). The maximum number of 17 alleles was detected at the locus CncirE2 in a collection specific to Andaman and Nicobar Islands (Rajesh et al., 2008b). Mean and range of PIC values of SSR loci (0.52, 0.14–0.78) obtained in the present study was also comparable with the previous studies. Rajesh et al. (2008a) reported that the PIC value ranged from 0.41 to 0.89 among 14 SSR loci with an average of 0.61. Ribeiro et al. (2010) reported that the PIC value ranged from 0.04 to 0.67 with a mean value of 0.44 among 13 SSR loci. In this study, the average gene diversity ( $H_e$ ) ranged from 0.16 to 0.81 with an average of 0.57. Similar levels of  $H_e$  estimates (range 0.42–0.66) were reported from other coconut germplasm collections in Brazil (Ribeiro et al., 2010), Kerala and Pondicherry region (Deva Kumar et al., 2011), Andaman and Nicobar Islands (Rajesh et al., 2008b), Sri Lanka (Perera et al., 2000) and France (CIRAD collection) (Gunn et al., 2011).

It is likely that small sample size as well as genotyping of only a single palm per accession would have resulted in underestimation of SSR allelic variation in this study. In addition, some alleles would have been missed by the gel system (6% PAGE), which can resolve only the alleles of at least 5–10 bp differences. However, our main interest was to assess the global picture of the diversity within a collection of world-wide coconut accessions (maintained at our location) and not within each accession. We also observed that overall SSR allelic variation in the collection was comparable with the published reports, which have used varied number of palms per accession (Perera et al., 2000, 2001; Meerow et al., 2003; Martinez et al., 2010). Overall, the results suggested that genetic diversity in coconut germplasm is moderately high, which might perhaps be due to its outcrossing nature and wide distribution. Hamrick and Godt

(1996) demonstrated that outcrossing species with long geographical range show more genetic diversity.

Cluster analysis divided the 79 coconut genotypes into two distinct groups with two major sub-clusters in each group indicating clear genetic structuring in the collection. The first cluster included most of the genotypes belonging to the Indo-Atlantic region comprising of Africa, Caribbean islands, Indian subcontinent and South Asia. Within this cluster, the ecotypes of Tamil Nadu, mostly collected from the farmer's field (>60 years old) in the southern districts, and Sri Lankan ecotypes were placed in a separate sub-cluster. These results were in agreement with Lebrun et al. (1998) who reported that the West African ecotypes were related to the Indian and Sri Lankan ecotypes. This clustering also supported the hypothesis of spread of coconut along the Atlantic Coasts of Africa and to the Panama region through nuts originating from the Indian Ocean.

The second group represented the Indo-Pacific Ocean basin which included all Talls belonging to the Indian Ocean, Oceania, South-East Asia and Pacific Islands along with the Dwarfs. This type of clustering pattern supports the hypothesis that Dwarf varieties could have originated from South Pacific or the South-East Asian region (Lebrun et al., 1999 and Gunn et al., 2011). Manimekalai and Nagarajan (2006) also observed the same relationship that Laccadive Dwarf and Chowghat Orange Dwarf from India representing South Asia, clustered with South Pacific Talls. Similarly, Perera et al. (2000) reported that Sri Lankan Dwarfs did not group with Sri Lankan Talls, but grouped with South-East Asian Tall coconuts. Lebrun et al. (1998) reported that African Dwarfs clustered with Far East and Pacific Tall coconuts. Similar to the findings of the current study, previous studies have also observed that South Asian and South-East Asian Talls such as Spicata and San Ramon, shared the cluster along with Dwarf genotypes suggesting that these Tall genotypes could have a common putative origin similar to Dwarfs from South-East Asia (Lebrun et al., 1998; Parthasarathy et al., 2004 and Dasanayaka et al., 2009). In this study, the green Dwarfs grouped with the Talls in the South-East Asia and Pacific group while the orange- and yellow-coloured Dwarfs grouped separately irrespective of their geographic origin. Using RAPD markers, Daher et al. (2002) and Upadhyay et al. (2004) also found that Malayan Dwarf genotypes clustered with the Brazilian and Indian Dwarf genotypes, respectively.

Although the dendrogram depicted a clear distinction between the Indo-Atlantic and the Indo-Pacific Ocean lineages of coconut genotypes, a few genotypes were mixed in both the groups. This suggested that the Indian Ocean region could have been a common point for the westward and eastward spread of coconut. These genotypes could have been dispersed either through human activity or could represent admixtures due to the outcrossing

nature. For instance, Jamaica Tall, a native of the Caribbean region and Nigerian Tall of Africa clustered with the South Asian genotypes particularly of Indian origin, which was in agreement with Manimekalai and Nagarajan (2006). Similarly, while most genotypes from the Philippines region clustered with South-East Asia and South Pacific group as expected, the Philippines Ordinary Tall genotypes with variations in the nut colour clustered with the Indo-Atlantic genotypes. Although dispersal through human activity cannot be ruled out, these results need to be interpreted with caution. It is also possible that these genotypes could indicate a hybrid or admixture from the native population due to outcrossing when domesticated over a period of time along with other exotic genotypes as part of the germplasm conservation programme.

The performances of the hierarchical clustering technique in germplasm collections have been proven with phenotypic as well as molecular data (Franco et al., 2005; Odong et al., 2011). In the hierarchical clustering technique, UPGMA (Sokal and Michener, 1958) and Ward (Ward, 1963) are the two most commonly used methods. While both the methods use Lance and William's recurrence formula (Lance and Williams, 1967), the algorithms differ in the method of defining the pairwise distance between clusters. In the present study, initially the UPGMA method was tried but it resulted in several clusters with one to few genotypes in each cluster (data not shown). Hence, the Ward's method of hierarchical clustering was employed, as it tends to form balanced clusters that could include the outlying genotypes (Jobson, 1992). The dendrogram thus obtained, formed distinct clusters which matched with groups based on geographic origin. This result was in good agreement with Odong et al. (2011). They compared both the UPGMA and Ward methods of hierarchical clustering in coconut and found that UPGMA placed outlying genotypes in separate clusters, which resulted in highly unbalanced clusters with many clusters containing one or two genotypes, and the clusters did not correspond with the geographic origin or passport data. The results of the current study also supported that Ward's method of hierarchical clustering would depict genetic relationships in coconut more realistically.

Presence of genetic structuring in coconut germplasm collection used in this study was further evident from the results of STRUCTURE analysis. Broadly, the collection was divided into two distinct populations ( $K=2$ ) namely the Indo-Atlantic Oceanic basins and Indo-Pacific Oceanic basins. Further sub-structuring was also clearly evident at  $K=4$ . These groupings were quite similar to dendrogram obtained by the Ward's method of hierarchical clustering. Most of the genotypes from Indo-Atlantic and South Asia regions were found in P2 and P4. Similarly, most of the genotypes from Indo-Pacific and South-East Asia regions were found in P1, P2 and P3. Considerable

number of genotypes (34) were admixture types. Several authors have documented population structure in different coconut germplasm collections namely France (CIRAD) (Gunn et al., 2011), Florida (Meerow et al., 2003; Mauro-Herrera et al., 2007), Brazil (Ribeiro et al., 2010; Loiola et al., 2016) and China (Xiao et al., 2013). It is reported that population structure in a species is affected by several factors including breeding systems, genetic drift, population size, seed dispersal, gene flow, evolutionary history and natural selection (Hamrick and Godt, 1990). In general, outcrossing species tend to show less structure compared with selfing species. On the contrary, clear structuring was observed in the coconut collection used in this study. Therefore, the results must be viewed in light of the following aspects: (1) some level of population structure is expected in coconut because both modes of reproduction namely cross pollination and self-pollination exist, (2) most of the coconut germplasm collections hitherto analyzed for structure have been artificially established collections through introduction and they were not naturally occurring populations; therefore, chances for gene flow would have been limited. For instances, in Florida and Brazil, coconut is not a native species and it has been established through introductions; expectedly, these collections showed more population differentiation and (3) the small population size would have affected the structure results.

Extent of LD (non-random association of alleles at different loci) is an important characteristic of a germplasm collection or a natural population for genetic studies. The results showed that the level of LD between unlinked SSR locus-pairs was very low, which is expected because the SSR loci were chosen based on genetic map position carefully avoiding closely linked markers. Small percentage of unlinked loci were in LD, which might be caused by population structure, mating system, co-selection of allele combinations at different loci, etc. (Hartl and Clark, 1997). However, it may be noted that LD observed in the current study would have been underestimated due to small number of marker loci and small population size. Nevertheless, low level of LD between unlinked loci is a desirable feature of the current germplasm panel for its use in association mapping studies.

Association analysis showed that a single SSR locus CnCir73 had putative association with several traits namely fruit breadth, nut weight, kernel weight and copra content explaining high phenotypic variance ( $R^2 = 34\text{--}43\%$ ). The association of CnCir73 with many fruit component traits is expected because they are highly correlated (Geethanjali et al., 2014). Interestingly, the marker CnCir73 was also detected in a quantitative trait loci (QTL) mapping study for fruit component traits by Baudouin et al. (2006). This QTL on chromosome 1 within the marker interval 37/2-CnCir73 was reported to be associated with fruit weight, nut weight, shell weight and meat nut ratio (Baudouin et al.,

2006). Thus, the results of the current study also support the association of this genomic region with fruit yield component traits in coconut, which need to be ascertained by further studies. To date, studies on QTL mapping are very limited in coconut (Herrán et al., 2000; Lebrun et al., 2001; Baudouin et al., 2006; Riedel et al., 2009) due to the great difficulty in developing mapping populations. Correspondence of the results from both association analysis and QTL mapping indicate that association mapping could be a useful alternative strategy for trait mapping in coconut.

However, it is important to note that the results of association analysis reported in this study is only preliminary and not conclusive due to small population size, which could lead to spurious associations. The germplasm panel also exhibited strong genetic structuring. Ideally, for a robust association study, an unstructured germplasm panel consisting of a large number of individuals must be genotyped at marker loci covering the whole genome and the phenotypic evaluation must be carried out at multiple locations. The use of large number of individuals in coconut has been a limitation due to perennial nature of the species and it would also be logistically demanding. In order to obtain reliable results, the population structure issue was addressed with the use of  $Q$  values as covariate in the model while performing the analysis. The FDR correction ( $q$ -value) of observed  $P$  values was also used to reduce the number of false-positive associations as suggested by Storey and Tibshirani (2003). In a similar study, Xiao et al. (2013) reported two SSR markers associated with height in a structured population containing 70 coconut accessions. A thorough genome-wide association study in a large germplasm panel is expected to provide more insights on genetic marker loci associations with phenotypic traits in coconut.

In conclusion, the present study revealed moderately high level of genetic diversity present in a coconut germplasm collection (79 genotypes) being maintained at TNAU, Coimbatore, one of the major coconut breeding centres in India, using a set of 48 SSR marker loci. The distance-based cluster analysis and the model-based structure analysis showed that the germplasm collection is genetically structured with two distinct lineages namely Indo-Atlantic-South Asia and Indo-Pacific-South-East Asia regions and further sub-divisions according to the geographical origin. The LD between unlinked SSR marker loci was very low (2.4%), which is a desirable feature for exploitation of this germplasm panel for association studies. In a preliminary association study, a single SSR locus CnCir73 on chromosome 1 was found to be putatively associated with fruit size and weight-related traits explaining high phenotypic variance (34–43%). Interestingly, this locus co-localized with a previously mapped QTL for fruit weight, nut weight, shell weight and meat nut ratio in coconut.

## Supplementary Material

The supplementary material for this article can be found at <https://doi.org/10.1017/S1479262117000119>.

## Acknowledgements

The authors thank Dr Senthilvel Senapathy, ICAR-IIOR for critical inputs on data analysis and Department of Science and Technology, Government of India for financial support through research grant No. SR/FT/LS-69/2010.

## References

- Abdurakhmonov IY and Abdurakimov A (2008) Application of association mapping to understanding the genetic diversity of plant germplasm resources. *International Journal of Plant Genomics* 2008: 574927. doi: 10.1155/2008/574927.
- Agrama HA, Eizenga GC and Yan W (2007) Association mapping of yield and its components in rice cultivars. *Molecular Breeding* 19: 341–356.
- Angeles JGC, Laurena AC and Tecson-Mendoza EM (2005) Extraction of genomic DNA from the lipid, polysaccharide and polyphenol-rich coconut (*Cocos nucifera* L.). *Plant Molecular Biology Reporter* 23(3): 297–298.
- Ashburner GR, Thompson WK and Halloran GM (1997) RAPD analysis of south Pacific coconut palm populations. *Crop Science* 37: 992–997.
- Balakrishnan NP and Nair RB (1979) Wild populations of *Areca* and *Cocos* in Andaman and Nicobar Islands. *Indian Journal of Forestry* 2: 350–363.
- Baudouin L and Lebrun P (2002) The development of microsatellite kit and dedicated software for use with coconuts. *BIOTROP Bulletin* 17: 16–20.
- Baudouin L, Lebrun P, Konan JL, Ritter E, Berger A and Billotte N (2006) QTL analysis of fruit components in the progeny of a Rennell Island Tall coconut (*Cocos nucifera* L.) individual. *Theoretical and Applied Genetics* 112: 258–268.
- Bradbury PJ, Zhang Z, Kroon DE, Casstevens TM, Ramdoss Y and Buckler ES (2007) TASSEL: software for association mapping of complex traits in diverse samples. *Bioinformatics* 23: 2633–2635.
- Cardena R, Ashburner GR and Oropeza C (2003) Identification of RAPDs associated with resistance to lethal yellowing of the coconut (*Cocos nucifera* L.) palm. *Scientia Horticulturae* 98: 257–263.
- Daher RF, Pereira MG, Tupinambá EA, Amaral Junior AT, Aragao WM, Ribeiro FE, Oliveira LO and Sakiyama NS (2002) Assessment of coconut tree genetic divergence by compound sample RAPD marker analysis. *Crop Breeding and Applied Biotechnology* 2(3): 431–438.
- Dasanayaka PN, Everard JMDT, Karunanayaka EH and Nandadasa HG (2009) Analysis of coconut (*Cocos nucifera* L.) diversity using microsatellite markers with emphasis on management and utilization of genetic resources. *Journal of the National Science Foundation of Sri Lanka* 37(2): 99–109.
- Deva Kumar K, Thomas RJ, Nair RV, Jerard AB, Rajesh MK, Jacob PM, Jayadev K and Parthasarathy VA (2011) Analysis of population structure and genetic relatedness among root (wilt) disease-resistant and susceptible west coast Tall coconut palms (*Cocos nucifera*) using microsatellite

- markers. *The Indian Journal of Agricultural Sciences* 81(6): 487–493.
- Earl DA and vonHoldt BM (2012) STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conservation Genetics Resources* 4(2): 359–361.
- Evanno G, Regnaut S and Goudet J (2005) Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Molecular Ecology* 14: 2611–2620.
- FAOSTAT (2012) <http://faostat.fao.org>.
- Font i Forcada C, Oraguzie N, Reyes-Chinwo S, Espiau MT, Socias i Company R and Fernández i Martí A (2015) Identification of genetic loci associated with quality traits in almond *via* association mapping. *PLoS ONE* 10(6): e0127656. doi: 10.1371/journal.pone.0127656.
- Franco J, Crossa J, Taba S and Shands H (2005) A sampling strategy for conserving genetic diversity when forming core subsets. *Crop Science* 45: 1035–1044.
- Geethanjali S, Rajkumar D and Shoba N (2014) Correlation and path coefficient analysis in coconut (*Cocos nucifera* L.). *Electronic Journal of Plant Breeding* 5(4): 702–707.
- Gunn BF, Baudouin L and Olsen KM (2011) Independent origins of cultivated coconut (*Cocos nucifera* L.) in the old world tropics. *PLoS ONE* 6(6): e21143 doi: 10.1371/journal.pone.0021143.
- Hamrick JL and Godt MJW (1990) *Plant Population Genetics, Breeding, and Genetic Resources*. Sunderland, MA: Sinauer. pp. 43–63.
- Hamrick JL and Godt MJW (1996) Effects of life history traits on genetic diversity in plant species. *Philosophical Transactions of Royal Society of London B* 351: 1291–1298.
- Harries HC (1978) The evolution, dissemination and classification of *Cocos nucifera* L. *The Botanical Review* 44: 265–320.
- Hartl DL and Clark AG (1997) *Principles of Population Genetics*. Sunderland, MA: Sinauer.
- Herrán A, Estioko L, Becker D, Rodriguez MJB, Rohde W and Ritter E (2000) Linkage mapping and QTL analysis in coconut (*Cocos nucifera* L.). *Theoretical and Applied Genetics* 101: 292–300.
- Iquiria E, Humira S and François B (2015) Association mapping of QTLs for Sclerotinia stem rot resistance in a collection of soybean plant introductions using a genotyping by sequencing (GBS) approach. *BMC Plant Biology* 15. DOI 10.1186/s12870-014-0408-y.
- Jobson JD (1992) *Applied Multivariate Data Analysis*, vol. II. Categorical and multivariate methods. New York: Springer.
- Lance GN and Williams WT (1967) A general theory of classificatory sorting strategies 1. Hierarchical systems. *The Computer Journal* 9: 373–380.
- Lebrun P, N'cho YP, Seguin M, Grivet L and Baudouin L (1998) Genetic diversity in coconut (*Cocos nucifera* L.) revealed by restriction fragment length polymorphism (RFLP) markers. *Euphytica* 101: 103–108.
- Lebrun P, Grivet L and Baudouin L (1999) Use of RFLP markers to study the diversity of the coconut palm. In: Oropeza C, Verdeil JL, Ashburner GR, Cardeña R, Santamaría JM (eds) *Current Advances in Coconut Biotechnology*. Netherlands: Springer, pp. 73–87.
- Lebrun P, Baudouin L, Bourdeix R, Konan JL, Barker JH, Aldam C, Herrán A and Ritter E (2001) Construction of a linkage map of the Rennell Island Tall coconut type (*Cocos nucifera* L.) and QTL analysis for yield characters. *Genome* 44: 962–970.
- Liu K and Muse SV (2005) PowerMarker-an integrated analysis environment for genetic marker analysis. *Bioinformatics* 21(9): 2128–2129.
- Liu X, Tang H, Li D and Hou L (2011) Genetic diversity of coconut cultivars in China by microsatellite (SSR) markers. *Molecular Plant Breeding* 2(12): 83–91.
- Loiola CM, Azevedo AON, Diniz LEC, Aragão WM, Azevedo CDDO, Santos PHAD, Ramos HCC, Pereira MG and Ramos SRR (2016) Genetic relationships among Tall coconut palm (*Cocos nucifera* L.) genotypes of the International Coconut Genebank for Latin America and the Caribbean (ICG-LAC), evaluated using microsatellite markers (SSRs). *PLoS ONE* 11(3): e0151309. doi: 10.1371/journal.pone.0151309.
- Manimekalai R and Nagarajan P (2006) Assessing genetic relationships among coconut (*Cocos nucifera* L.) genotypes using inter simple sequence repeat markers. *Scientia Horticulturae* 108: 49–54.
- Martinez RT, Baudouin L, Berger A and Dollet M (2010) Characterization of the genetic diversity of the Tall coconut (*Cocos nucifera* L.) in the Dominican Republic using microsatellite (SSR) markers. *Tree Genetics & Genomes* 6: 73–81.
- Mauro-Herrera M, Meerow AW, Borrone JW, Kuhn DN and Schnell RJ (2007) Usefulness of WRKY gene-derived markers for assessing genetic population structure: an example with Florida coconut cultivars. *Scientia Horticulturae* 115: 19–26.
- Meerow AW, Wissner RJ, Brown JS, Kuhn DN, Schnell RJ and Broschat TK (2003) Analysis of genetic diversity and population structure within Florida coconut (*Cocos nucifera* L.) germplasm using microsatellite DNA, with special emphasis on the Fiji Dwarf cultivar. *Theoretical and Applied Genetics* 106: 715–726.
- Odong TL, van Heerwaarden J, Jansen J, van Hintum TJJ and van Eeuwijk FA (2011) Determination of genetic structure of germplasm collections: are traditional hierarchical clustering methods appropriate for molecular marker data? *Theoretical and Applied Genetics* 123: 195–205.
- Panaud O, Chen X and McCouch SR (1996) Development of microsatellite markers and characterization of simple sequence length polymorphism (SSLP) in rice (*Oryza sativa* L.). *Molecular and General Genetics* 252: 597–607.
- Panda KC (1982) Coconut types of Orissa. *Indian Coconut Journal* 13: 10–18.
- Parthasarathy VA, Geethalakshmi P and Niral V (2004) Analysis of coconut cultivars and hybrids using isozyme polymorphism. *Acta Botanica Croatica* 63: 69–74.
- Perera L, Russell JR, Provan J, McNicol JW and Powell W (1998) Evaluating genetic relationship between indigenous coconut (*Cocos nucifera* L.) genotypes from Sri Lanka by means of AFLP profiling. *Theoretical and Applied Genetics* 96: 545–550.
- Perera L, Russell JR, Provan J and Powell W (2000) Use of microsatellite DNA markers to investigate the level of genetic diversity and population genetic structure of coconut (*Cocos nucifera* L.). *Genome* 43: 15–21.
- Perera L, Russell JR, Provan J and Powell W (2001) Levels and distribution of genetic diversity of coconut (*Cocos nucifera* L., var. *Typica* form *typica*) from Sri Lanka assessed by microsatellite markers. *Euphytica* 122: 381–389.
- Perrier X and Jacquemoud-Collet JP (2006) DARwin software. Available at <http://darwin.cirad.fr/>
- Pritchard JK, Stephens M, Rosenberg NA and Donnelly P (2000) Association mapping in structured populations. *The American Journal of Human Genetics* 37: 170–181.
- Rajesh MK, Arunachalam V, Nagarajan P, Lebrun P, Samsudeen K and Thamban C (2008a) Genetic survey of 10 Indian coconut landraces by simple sequence repeats (SSRs). *Scientia Horticulturae* 118: 282–287.

- Rajesh MK, Nagarajan P, Jerard BA, Arunachalam V and Dhanapal R (2008b) Microsatellite variability of coconut genotypes (*Cocos nucifera* L.) from Andaman and Nicobar Islands. *Current Science* 94: 12.
- Ratnambal MJ, Nair MK, Muralidharan K, Kumaran PM, Bhaskara Rao EVV and Pillai RV (1995) *Coconut Descriptors. Part I*. Kasaragod, India: Central Plantation Crops Research Institute.
- Ribeiro FE, Baudouin L, Lebrun P, Chaves LJ, Brondani C, Zucchi MI and Vencovsky R (2010) Population structures of Brazilian Tall coconut (*Cocos nucifera* L.) by microsatellite markers. *Genetics and Molecular Biology* 33(4): 696–702.
- Riedel M, Riederer M, Becker D, Herrán A, Kullaya A, Arana-López G, Peña-Rodríguez L, Billotte N, Sniady V, Rohde W and Ritter E (2009) Cuticular wax composition in *Cocos nucifera* L.: physicochemical analysis of wax components and mapping of their QTLs onto the coconut molecular linkage map. *Tree Genetics & Genomes* 5: 53–69.
- Rivera R, Edwards KJ, Barker JH, Arnold GM, Ayad G, Hodgkin T and Karp A (1999) Isolation and characterization of polymorphic microsatellites in *Cocos nucifera* L. *Genome* 42: 668–675.
- Shalini KV, Manjunatha S, Lebrun P, Berger A, Baudouin L, Pirany N, Ranganath RM and Prasad DT (2007) Identification of molecular markers associated with mite resistance in coconut (*Cocos nucifera* L.). *Genome* 50: 35–42.
- Sokal RR and Michener CD (1958) A statistical method for evaluating systematic relationships. *University of Kansas Science Bulletin* 38: 1409–1438.
- Storey JD and Tibshirani R (2003) Statistical significance for genome wide studies. *Proceedings of National Academy of Sciences USA* 100(16): 9440–9445.
- Sugimura Y, Itano U, Salud CD, Otsuji K and Yamaguchi H (1997) Biometric analysis of diversity of coconut palm: cultivar classification by botanical and agronomical traits. *Euphytica* 98: 29–35.
- Tautz D (1989) Hypervariability of simple sequences as a general source of polymorphic DNA markers. *Nucleic Acids Research* 17: 6463–6471.
- Teulat B, Aldam C, Trehin R, Lebrun P, Barker JHA, Arnold GM, Karp A, Baudouin L and 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. *Theoretical and Applied Genetics* 100: 764–771.
- Upadhyay A, Jayadev K, Manimekalai R and Parthasarathy VA (2004) Genetic relationship and diversity in Indian coconut genotypes based on RAPD markers. *Scientia Horticulturae* 99: 353–362.
- Ward JH (1963) Hierarchical groupings to optimize an objective function. *Journal of the American Statistical Association* 58: 236–244.
- Xiao Y, Luo Y, Yang Y, Fan H, Xia W, Mason AS, Zhao S, Sager R and Qiao F (2013) Development of microsatellite markers in *Cocos nucifera* and their application in evaluating the level of genetic diversity of *Cocos nucifera*. *Plant Omics Journal* 6(3): 193–200.
- Xia W, Xiao Y, Liu Z, Luo Y, Mason AS, Fan H, Yang Y, Zhao S and Peng M (2014) Development of gene-based simple sequence repeat markers for association analysis in *Cocos nucifera*. *Molecular Breeding* 34: 525–535.
- Zizumbo-Villarreal D and Piñero D (1998) Pattern of morphological variation and diversity of *Cocos nucifera* (Arecaceae) in Mexico. *American Journal of Botany* 85(6): 855–865.