



# Genetic characterization of coconut (*Cocos nucifera* L.) varieties conserved in Vietnam through SCoT marker-based polymorphisms

Nguyen Hoang Minh Khang · Nguyen Thien Quang · Hoang Ngoc Xuan Mai ·  
Nguyen Doan Nguyen Phuong · Nguyen Phuong Thao · Nguyen Bao Quoc

Received: 18 December 2020 / Accepted: 19 June 2021  
© The Author(s), under exclusive licence to Springer Nature B.V. 2021

**Abstract** Coconut (*Cocos nucifera* L.) production sustains millions of smallholder farmers and their families around the globe, and also powers a thriving industry of climate-adaptive nutrition and biomaterials. Demand for high-quality planting materials has risen sharply in recent years, calling for greater understanding and potential utilization of the genetic diversity in coconut populations. Genetic diversity among several conserved coconut varieties in Vietnam was investigated to assess the suitability of 15 start codon targeted (SCoT) primers for future selective breeding programs. Amplification of 15 SCoT primers in 57 individuals from 19 coconut varieties revealed high intra-variety diversity (Nei's genetic diversity index = 0.237, Shannon's Information index = 0.352), especially among the Vietnamese Dwarfs.

Intriguingly, inter-variety differentiation accounted for only 9.41% of total genetic variation. The designated SCoT primers were moderately informative. On average, each primer had 9.5 reproducible polymorphic bands per reaction. Polymorphism Information Content (PIC) per locus ranged between 0.217 and 0.390, with Resolving power (Rp) between 0.271 and 0.599. Principal Coordinate Analysis (PCoA) and Unweighted Pair Group Method with Arithmetic mean (UPGMA) analyses revealed distinct clusters for non-native Dwarf, native Dwarf, and Tall varieties. This paper represents the first-of-its-kind attempt to shed light on coconut genetic diversity in Vietnam using molecular markers.

**Keywords** Coconut · SCoT · Genetic diversity · Dwarf · Tall · Varieties

N. H. M. Khang · N. T. Quang · N. P. Thao (✉)  
Applied Biotechnology for Crop Development Research  
Unit, School of Biotechnology, International University,  
Quarter 6, Linh Trung Ward, Thu Duc,  
Ho Chi Minh 700000, Vietnam  
e-mail: npthao@hcmiu.edu.vn

N. H. M. Khang · N. T. Quang · N. P. Thao  
Vietnam National University, Linh Trung Ward, Thu Duc,  
Ho Chi Minh 700000, Vietnam

H. N. X. Mai · N. D. N. Phuong · N. B. Quoc (✉)  
Research Institute for Biotechnology and Environment,  
Nong Lam University, Linh Trung ward, Thu Duc,  
Ho Chi Minh, Vietnam  
e-mail: baoquoc@hcmuaf.edu.vn

## Abbreviations

<i>SCoT</i>	Start codon targeted
<i>QTL</i>	Quantitative trait locus
<i>PIC</i>	Polymorphism information content
<i>Rp</i>	Resolving power
<i>PCoA</i>	Principal coordinate analysis
<i>UPGMA</i>	Unweighted pair group method with arithmetic mean

## Introduction

Coconut (*Cocos nucifera* L.) has long been appreciated worldwide as a multi-functional crop with economic, environmental and social significance (Niral and Jerard 2018). In Vietnam, commercialization of fresh coconuts and related products accounts for a significant percentage of household income in productive provinces, particularly in Ben Tre, Tien Giang, and Tra Vinh. Following molecular techniques developed for important crops such as rice (*Oryza sativa* L.) and corn (*Zea mays* L.), the coconut genetic diversity has increasingly been studied and characterized (Rajesh et al. 2018). In the past, diversity studies primarily related to the presence of polyphenols, isozymes, carotenols or protein markers. Amongst these early markers, analysis of polyphenols using high-performance liquid chromatography revealed the highest levels of polymorphism (Chempakam and Ratnambal 1993). Since the 1990s, DNA markers have been widely used to analyze coconut genetic diversity, with copious studies using Random Amplification of Polymorphic DNA (RAPD) and Simple Sequence Repeats (SSR) techniques (Ashburner et al. 1997; Everard et al. 1996; Hayati and Hartana 2000; Rivera et al. 1999; Perera et al. 2000; Meerow et al. 2003; Gunn et al. 2011; Loiola et al. 2016; Preethi et al. 2020). Earlier studies used other marker systems including Restriction Fragment Length Polymorphism (RFLP) (Lebrun et al. 1998), Inverse Sequence-Tagged Repeat (ISTR) (Rohde et al. 1996; Duran et al. 1997), Inter Simple Sequence Repeats (ISSR) (Manimekalai et al. 2006), WRKY Loci (Mauro-Herrera et al. 2007; Meerow et al. 2009) and Start Codon Targeted Polymorphism (SCoT) (Rajesh et al. 2015). Major achievements in coconut genetic study have successfully contributed to (i) differentiation based on the level of genetic variation in Tall and Dwarf varieties (Perera et al. 2000), (ii) construction of genetic maps (Rohde et al. 2002; Lebrun et al. 2001), (iii) construction of phylogenetic dendrograms providing evidence for origin and distribution (Upadhyya et al. 2004; Dasanayaka et al. 2009), (iv) suggested qualitative trait loci (QTLs) for mite resistance and the pandan aroma (Shalini et al. 2007; Saensuk et al. 2016), and (v) identification of genes involved in somatic embryogenesis, lipid biosynthesis and salt tolerance (Rajesh et al. 2016a; Reynolds et al. 2019; Yang et al. 2021).

Start codon targeted (SCoT) polymorphism is a straightforward and powerful DNA marker system that was first presented by Collard and Mackill (2009). The single, random primers used in PCR reactions are based on the short conserved region surrounding the ATG translation start codon in plant genes on both DNA strands. Point mutations at the primer-binding site or indel mutations within the amplified regions contribute to varying banding patterns that have successfully been used to characterize polymorphism in rice (Collard and Mackill 2009), longan (Chen et al. 2010) and grapes (Zhang et al. 2011). A screening study by Rajesh et al. (2015), involving 25 markers as previously described by Collard and Mackill (2009), found that 15 showed consistent amplification patterns for a set of 23 coconut varieties, and were therefore recommended for further analysis. SCoT markers have also proved useful in identifying true hybrid seedlings in areca nut palms (*Areca catechu* L.), a member of the same Areca family as coconut. Based on a 1300 bp polymorphic band produced by primer SCoT11, a sequence-characterized amplified region (SCAR) marker, which can distinguish between true Areca nut Tall x Hirehalli Dwarf hybrid seedlings and false seedlings produced due to self-pollination of Hirehalli Dwarf mother palms was developed (Rajesh et al. 2016b). Here, the suitability of 15 SCoT primers was examined for future selective breeding programs in Vietnamese coconuts, while genetic diversity of a subset of varieties conserved at the Dong Go Coconut Research Center (Ben Tre, Vietnam) was also examined.

## Materials and methods

### Plant materials and DNA isolation

Genetic materials were sourced from 19 coconut varieties, comprising thirteen Dwarf and six Tall varieties (Table 1). Each variety was represented by three different individual palms. Fresh leaves were thoroughly collected from the Dong Go Coconut Research Centre (Giong Trom, Ben tre province, Vietnam). DNA samples were isolated from 57 samples using the extraction procedure of the GeneJet Plant Genome Extraction Kit (Thermo Scientific) as described by the manufacturer. The quality and quantity of DNA were verified using 1.0% agarose

**Table 1** List of coconut varieties used in the study

#	Variety	Dong Go abbreviation	International abbreviation	Water weight (g)	Oil content/dry weight (%)	Fruit height (cm)	Fruit width (cm)	Sugar content in 8-month-old fruit (°Bx)	No. of fruit/tree/year (average)
1	Ghana Yellow Dwarf	GYD	GYD	133.3	63.9	50	43.3	NA	110
2	Equatorial Guinea Green Dwarf	EGD	EGD	220	63.7	54.6	52.3	6.75	110
3	Catigan Green Dwarf	CAT	CAT	216.7	NA	55.6	54.7	6.75	110
4	Malayan Yellow Dwarf	MYD	MYD	153.4	63.7	NA	NA	6.75	145
5	Malayan Red Dwarf	MRD	MRD	282.5	68	56.6	54.7	6.75	110
6	Sri Lanka Red Dwarf	SLRD	SLRD	133.3	63.3	52	47.3	NA	110
7	Soc Dwarf <sup>a</sup>	SOCD	SOCD <sup>b</sup>	102.5	58	NA	NA	8	145
8	Eo Brown Dwarf <sup>a</sup>	EOBD	EOD	30.6	63.4	NA	NA	7.25	250
9	Tam Quan Yellow Dwarf <sup>a</sup>	TQYD	TQYD	357.3	49.4	50	43.6	7.25	90
10	Xiem Green Dwarf <sup>a</sup>	XGD	XGD	396	64	57.6	58.6	NA	110
11	Xiem Luc Green Dwarf <sup>a</sup>	XLGD	XLGD <sup>b</sup>	249.8	64.7	54	53	NA	70
12	Aromatic Green Dwarf <sup>a</sup>	AROGD	AROD	343.3	63.73	54.6	52.3	6.75	80
13	Xiem Fire Dwarf <sup>a</sup>	XFD	XFD <sup>b</sup>	336.7	61.99	58	58.3	NA	85
14	Dau Red Tall <sup>a</sup>	DAURT	DAURT <sup>b</sup>	250	64.45	53.4	44	NA	175
15	Macapuno Tall <sup>a</sup>	MACT	MACT02	505	65	66.4	64.1	NA	80
16	San Ramon Tall	SNRT	SNRT	113.1	60.7	46.9	39.6	7.25	145
17	Bago Oshiro Tall	BAOT	BAOT	127	66	36.8	42.8	8.5	145
18	West African Tall	WAT	WAT	122.5	58.7	44.3	21	8.25	145
19	Ta Green Tall <sup>a</sup>	TAGT	TAAT01	181	NA	41.8	23	7.25	145

<sup>a</sup>Varieties marked with <sup>a</sup> are local varieties from Vietnam

<sup>b</sup>Abbreviations for these varieties are not available in the Cogent Coconut Genetic Resources Database (CGRD), thus Dong Go abbreviations are used

gel electrophoresis, a NanoDrop Spectrophotometer and PCR amplification of the plant's housekeeping gene COX (mitochondrial cytochrome oxidase).

### SCoT markers analysis

Three biological replicates were used for each variety to examine and account for the within-population genetic diversity of each variety. All samples were amplified by 15 SCoT primers, previously screened for clarity by Rajesh et al. (2015). Details of the primers, synthesized from Integrated DNA Technologies (Vietnam) are provided in Table 2. The reaction mixture had a total volume of 25  $\mu$ L and contained 5  $\mu$ L 5  $\times$  MyTaq Buffer, 0.25  $\mu$ L MyTaq DNA Polymerase (Bioline), 2  $\mu$ L primer (10  $\mu$ M), 15.75  $\mu$ L distilled water and 2  $\mu$ L template DNA (3 ng/ $\mu$ L). Concentration of MyTaq DNA Polymerase in the reaction mixture was 0.05 units/ $\mu$ L. All PCR amplifications were carried out in a GeneQ<sup>TM</sup> Thermal Cycler (Bioer, China) with the following conditions: initial denaturation at 94  $^{\circ}$ C for 3 min, followed by 35 cycles of denaturation at 94  $^{\circ}$ C for 1 min, annealing for 1 min at variable melting temperature ( $T_m$ ) for each primer as presented in Table 2 and then 72  $^{\circ}$ C for 1 min. The final extension at 72  $^{\circ}$ C lasted 7 min. Three microliters of PCR products from each reaction were mixed with 1.5  $\mu$ L of 0.6  $\times$  GelRed<sup>TM</sup> (ABT,

Vietnam), and separated on 1.5% agarose gel in 0.5  $\times$  TBE buffer by electrophoresis at 50 V for 45 min using HyperLadder<sup>TM</sup> 1 kb (Bioline) as a molecular weight standard. Gels were visualized under UV light in a UVP<sup>TM</sup> GelDoc-It<sup>®</sup> imaging system.

### Data analysis

The gel images were analyzed by CLIQS 1D software, version 1.4 from TotalLab (Newcastle, UK). The bands were scored as either present (1) or absent (0), and their respective molecular weights were calculated in base pairs (bp). Band matching for each gel image was automatically carried out by the program. If the retardation factor (Rf) value of the bands was within 0.015 units of another, a match was declared.

The Polymorphic Information Content (PIC) of each locus was calculated using the formula for dominant markers as  $PIC = 1 - \sum_i f_i^2$ , where  $f_i$  is the frequency of the  $i$ th allele (De Riek et al. 2001). The Resolving power (Rp) of each locus was calculated according to a previous study as  $Rp = \sum Ib$ , where  $Ib = 1 - (2 \times |0.5 - f_i|)$  (Prevost and Wilkinson 1999). The heterozygosity (He) value of each variety was computed according to Nei's formula for within population variation index as  $He = 1 - \sum_{i=1}^n f_i^2$  (Nei

1978). The unbiased heterozygosity (uHe) value was calculated as  $uHe = [2N \div (2N - 1)] * He$  where N is the number of samples. The Shannon's Information index was determined using  $I = - \sum f_i \times \log_2(f_i)$  (Lewontin 1972). The Shannon's Information index (uI) value was calculated as  $uI = [2N \div (2N - 1)] * I$ .

Multi-dimensional scaling analysis was undertaken by applying the Principal Coordinate Analysis (PCoA) method to the dissimilarity matrix composed of populations' pairwise Nei's genetic distances. Analysis of MOlecular VAriance (AMOVA) was performed with 9999 permutations for all data. Hierarchical cluster analysis was achieved with the application of UPGMA to the matrix of Simple Matching similarity coefficients. Calculations and data visualizations were performed using Microsoft Excel (2010), Genalex 6.5 (Peakall and Smouse 2006, 2012), PopGene 1.32 (Yeh et al. 1997) and NTSYS-pc 2.1 (Rohlf et al. 2002).

**Table 2** Sequence and melting temperature of SCoT primers

STT	Primer	3'-5' sequence	$T_m$ ( $^{\circ}$ C)
1	SCOT01	CAACAATGGCTACCACCA	52.6
2	SCOT04	CAACAATGGCTACCACCT	52.3
3	SCOT05	CAACAATGGCTACCACGA	52.6
4	SCOT13	ACGACATGGCGACCATCG	58.0
5	SCOT14	ACGACATGGCGACCACGC	61.3
6	SCOT15	ACGACATGGCGACCGCGA	62.6
7	SCOT16	ACCATGGCTACCACCGAC	57.3
8	SCOT17	ACCATGGCTACCACCGAG	57.1
9	SCOT18	ACCATGGCTACCACCGCC	60.7
10	SCOT19	ACCATGGCTACCACCGGC	60.7
11	SCOT20	ACCATGGCTACCACCGCG	60.8
12	SCOT21	ACGACATGGCGACCCACA	59.7
13	SCOT22	AACCATGGCTACCACCAC	55.0
14	SCOT23	CACCATGGCTACCACCAG	55.7
15	SCOT25	ACCATGGCTACCACCGGG	60.2

## Results

### SCoT polymorphism

A total of 3774 amplicons were obtained, with *ca.* 251.6 bands per primer. The length of the bands ranged from 150 to 8500 bp. By matching molecular weights of the amplicons between corresponding replicates, we discovered 306 unique loci with an average of 20.4 unique resolvable bands amplified per primer. The number of loci with reproducible and polymorphic bands was 180, with an average of 9.5 bands per primer, ranging from 5 bands (SCOT19) to 17 bands (SCOT13). The  $R_p$  values ranged between 1.895 (SCOT04) and 9.263 (SCOT13), with an average of 5.733. Total Polymorphic Information Content (PIC) ranged from 1.518 (SCOT04) to 6.039 (SCOT13), with an average of 3.849. Average PIC over loci ( $PIC_{ave}$ ) ranged from 0.217 (SCOT04) to 0.390 (SCOT18) with an average of 0.316. Table 3 presents all data regarding number of amplicons and loci,  $R_p$  and PIC values for each SCoT primer. PCR amplification profiles, produced with two

representative SCOT primers (SCOT04 and SCOT15), were shown in Fig. 5.

### Genetic diversity analysis among coconut varieties

Based on allele distribution statistics presented in Table 4, the frequency of all reproducible polymorphic loci was greater than 5%, suggesting that no band was unique to a single variety. Numbers of bands were higher in XFD (158), AROD (151), XLGD (150) while MRD (29), EGD (37), and SLRD (45) were recorded the lowest number of bands. Furthermore, the number of bands presented at less than or equal to 25% of the population was observed in XFD (11) and in MYD, MRD, SLRD and TQYD (0). The number of less common bands presented at less than or equal to 50% of the populations was reported in XFD (50) and MRD (2). Likewise, the mean unbiased expected heterozygosity ( $uHe$ ) ranged from 0.058 (MRD) to 0.410 (XFD), with an average of 0.237. The mean unbiased Shannon's Information index ranged between 0.092 (MRD) and 0.592 (XFD), with an average of 0.352. Applying the two-tailed t-test and

**Table 3** Primer information, scored bands and discriminating power statistics

Primer	Na	Nu	Nr	$R_{p_s}$	$R_{p_{ave}}$	$PIC_s$	$PIC_{ave}$
SCOT01	275	23	9	4.526	0.503 ± 0.350	2.837	0.315 ± 0.176
SCOT04	117	9	7	1.895	0.271 ± 0.185	1.518	0.217 ± 0.122
SCOT05	259	15	9	4.421	0.491 ± 0.281	2.981	0.331 ± 0.139
SCOT13	358	19	17	9.263	0.545 ± 0.287	6.039	0.355 ± 0.138
SCOT14	256	19	15	6.000	0.400 ± 0.303	4.111	0.274 ± 0.157
SCOT15	360	32	15	8.211	0.547 ± 0.301	5.285	0.352 ± 0.145
SCOT16	242	22	16	8.105	0.507 ± 0.297	5.396	0.337 ± 0.141
SCOT17	248	21	13	5.158	0.397 ± 0.252	3.723	0.286 ± 0.125
SCOT18	288	27	13	7.789	0.599 ± 0.246	5.064	0.390 ± 0.111
SCOT19	183	20	5	2.526	0.505 ± 0.181	1.806	0.361 ± 0.074
SCOT20	219	18	11	4.211	0.383 ± 0.267	3.014	0.274 ± 0.131
SCOT21	292	23	14	7.579	0.541 ± 0.305	4.875	0.348 ± 0.153
SCOT22	260	21	12	6.316	0.526 ± 0.272	4.211	0.351 ± 0.135
SCOT23	262	19	14	7.053	0.504 ± 0.296	4.665	0.333 ± 0.156
SCOT25	155	18	10	2.947	0.295 ± 0.244	2.216	0.222 ± 0.125
Mean	251.6	20.4	9.5	5.733	0.486 ± 0.096	3.849	0.316 ± 0.051

*Na* Number of amplicons; *Nu* Number of unique loci ranging from 200 to 5000 bp in length; *Nr* Number of reproducible polymorphic loci;  $R_{p_s}$  Sum of  $R_p$  values of all reproducible polymorphic loci amplified by the primer;  $R_{p_{ave}}$  Average of PIC values of all reproducible polymorphic loci amplified by the primer;  $PIC_s$  Sum of PIC values of all reproducible polymorphic loci amplified by the primer;  $PIC_{ave}$  Average of PIC values of all reproducible polymorphic loci amplified by the primer

**Table 4** List of allele distribution statistics for 19 coconut varieties

#	1	2	3	4	5	6	7	8	9	14
V.ar. Abbr	GYD	EGD	CAT	MYD	MRD	SLRD	SOC	EOD	TQYD	DAURT
Nu	71	37	84	69	29	45	142	90	140	145
Nu 0.05	71	37	84	69	29	45	142	90	140	145
NL 0.25	6	1	2	0	0	0	1	3	0	4
NL 0.50	21	3	17	7	2	5	37	13	35	38
Mean uHe	0.18 ± 0.018	0.084 ± 0.013	0.181 ± 0.016	0.147 ± 0.015	0.058 ± 0.010	0.100 ± 0.013	0.334 ± 0.018	0.198 ± 0.016	0.324 ± 0.017	0.359 ± 0.017
Mean ul	0.264 ± 0.348	0.126 ± 0.262	0.275 ± 0.262	0.226 ± 0.303	0.092 ± 0.211	0.154 ± 0.271	0.488 ± 0.344	0.302 ± 0.321	0.483 ± 0.317	0.524 ± 0.329
#	15	16	17	18	19	10	11	12	13	
V.ar. Abbr	MACT02	SNRT	BAOT	WAT	TAAT01	XGD	XLGD	AROD	XFD	
Nu	96	133	114	69	85	127	150	151	158	
Nu 0.05	96	133	114	69	85	127	150	151	158	
NL 0.25	1	4	2	2	2	3	3	2	11	
NL 0.50	12	37	26	13	12	29	42	43	50	
Mean uHe	0.217 ± 0.017	0.288 ± 0.018	0.292 ± 0.019	0.166 ± 0.017	0.205 ± 0.017	0.278 ± 0.016	0.356 ± 0.018	0.335 ± 0.018	0.410 ± 0.016	
Mean ul	0.325 ± 0.34	0.423 ± 0.352	0.427 ± 0.358	0.248 ± 0.332	0.306 ± 0.341	0.422 ± 0.316	0.518 ± 0.338	0.490 ± 0.342	0.592 ± 0.306	

*Nu* Number of different bands; *Nu* 0.05 Number of different bands with frequency > = 5%; *NL* 0.25 Number of locally common bands with a frequency > = 5% found in < = 25% of the population; *NL* 0.50 Number of locally common bands with a frequency > = 5% found in < = 50% of the population; *Mean uHe* Average of unbiased expected heterozygosity; *Mean ul* Average of unbiased Shannon's Information index for a variety

assuming unequal variances, the Nei's genetic diversity index ( $uHe$ ) of Vietnamese Dwarfs ( $M = 0.319$ ,  $SD = 0.066$ ) was higher than the non-native Dwarf varieties ( $M = 0.260$ ,  $SD = 0.086$ ),  $t(11) = 5.916$ ,  $p < 0.001$ . AMOVA analysis presented in Table 5 indicated significant variation within varieties (90.57%,  $\Phi_{PT} = 0.094$ ,  $p = 0.009$ ).

#### Multi-dimensional scaling analysis

The first three axes of the PCoA plot derived from Nei's genetic distances represented 34.96%, 10.45% and 7.81% of the cumulative variation among varieties. The 19 coconut varieties were separated into four distinct groups, as shown in Fig. 1. The clearest association was seen among six non-native Dwarf varieties—one Sri Lankan (SLRD), two African (GYD, EGD), two Malaysian Dwarfs (MYD, MRD) and one Filipino Dwarf (CAT). Four Vietnamese varieties, including three Dwarfs and one hybrid (XLGD, AROD, XGD and DAURT) were also clustered together. The third group was made up of four Tall varieties (WAT, BAOT, TAAT01 and MACT02) and one Dwarf (EOD). Two Dwarfs from Vietnam primarily harvested for drinking water (SOCD and TQYD) exhibited a close relationship, according to this analytical tool. As data showed, XFD and SNRT showed no meaningful association with any group.

#### Hierarchical clustering analysis

The average pairwise Simple Matching (SM) coefficient was 0.611. The pairwise Simple Matching (SM) coefficient was lowest in two Dwarf varieties from Sri Lanka (SLRD) and Vietnam (XFD) at 0.411, and highest in two Dwarf varieties from Sri Lanka (SLRD) and Malaysia (MRD) at 0.956. The Vietnamese XFD showed the lowest coefficients against all foreign Dwarf varieties. Bootstrap values for branches on the

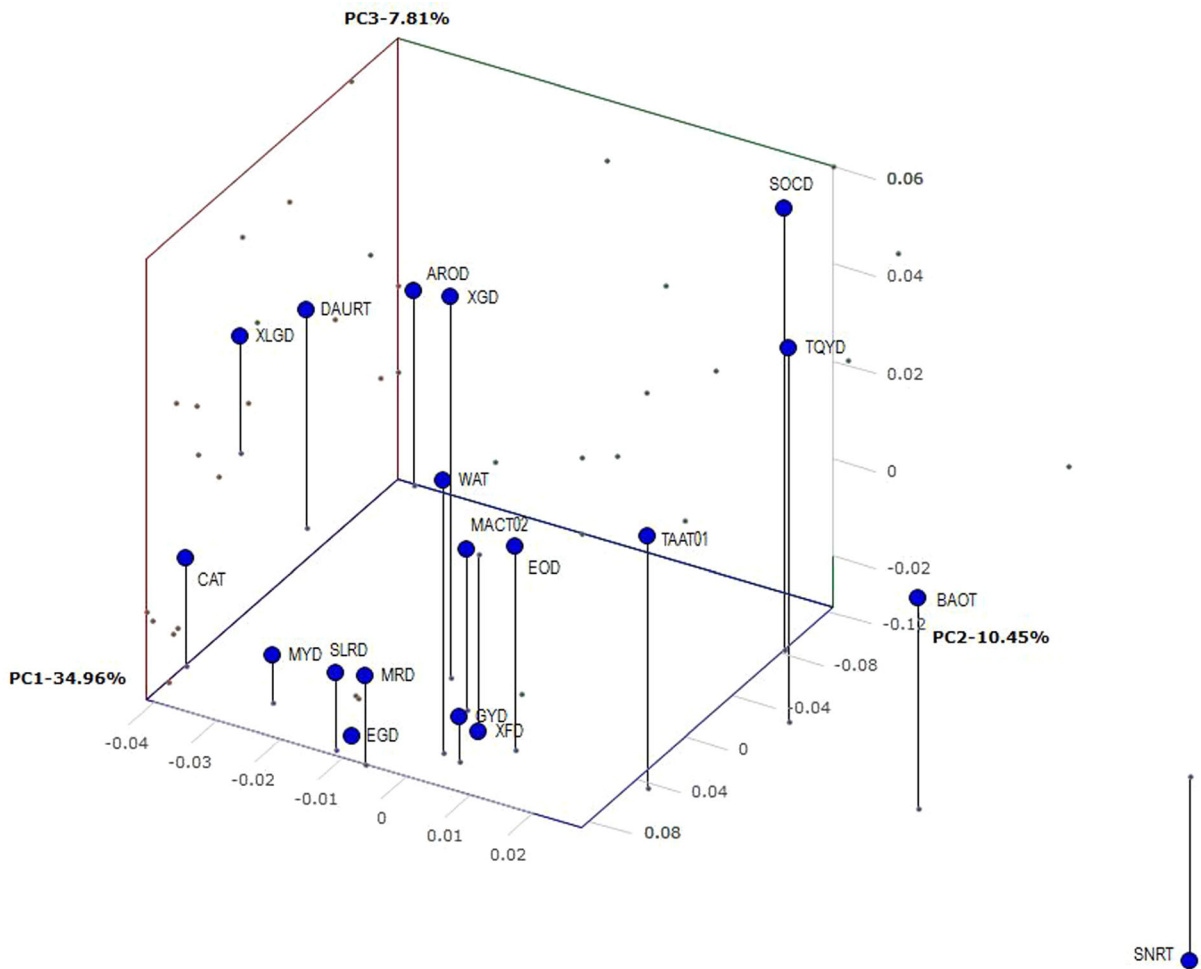
dendrogram constructed from UPGMA analysis ranged between 26 and 90%, as shown in Fig. 2. Two main clusters were identified, designated as 'A' and 'B' diverging at the similarity coefficient of 0.550, meaning that they shared 55% of the amplified bands. Each cluster contained two clades as A.1, A.2 and B.1, B.2. Cluster A was characterized by five varieties (DAURT, SOCD, XLGD, AROD and XFD) that all had small fruit sizes and high sugar levels of 7.5–8.5°Bx. Within Cluster A, Clade A.1 contained one hybrid (DAURT) and two Dwarf (SOCD, XLGD) varieties from Vietnam, while Clade A.2 contained two Dwarf varieties from Vietnam (AROD and XFD). Cluster B contained both Tall and Dwarf varieties. Sub-cluster B.1 contained two clades with the highest similarity values (Clade B.1.1 and Clade B.1.2), while Clade B.1.1 contained six Dwarf varieties from Equatorial Guinea, Malaysia, Sri Lanka, the Philippines, and Vietnam (EGD, MYD, MRD, SLRD, CAT, and EOD). These varieties produced coconut water at medium quantity (133.3–220 g per nut), with sugar level of 6.75°Bx (data not shown). Clade B.1.2 contained three Tall varieties from Vietnam and West Africa (MACT02, WAT and TAAT01. These all had oil content in the upper-medium range (*ca.* 65% of dry weight) (data not shown). Sub-cluster B.2 contained a Vietnamese Dwarf variety (TQYD) and two Tall varieties from the Philippines (SNRT and BAOT). These varieties were less productive in fruit yield (80–90 fruits/tree/year).

#### Discussion

The SCoT system was designed with longer primers, modeled after short conserved regions flanking ATG translation start codon in plants. This characteristic confers higher reproducibility, with a greater possibility of amplifying meaningful regions of the genome compared to RAPD markers (Collard and Mackill

**Table 5** Analysis of Molecular Variance (AMOVA) for 19 coconut varieties

Source of variation	Degree of freedom	Sum of squares	Mean squares	Estimated variation	Percentage variation
Among varieties	18	734.772	40.821	3.238	9.41%
Within varieties	38	1182.000	31.105	31.105	90.57%
Total	56	1916.772		34.344	100.00%

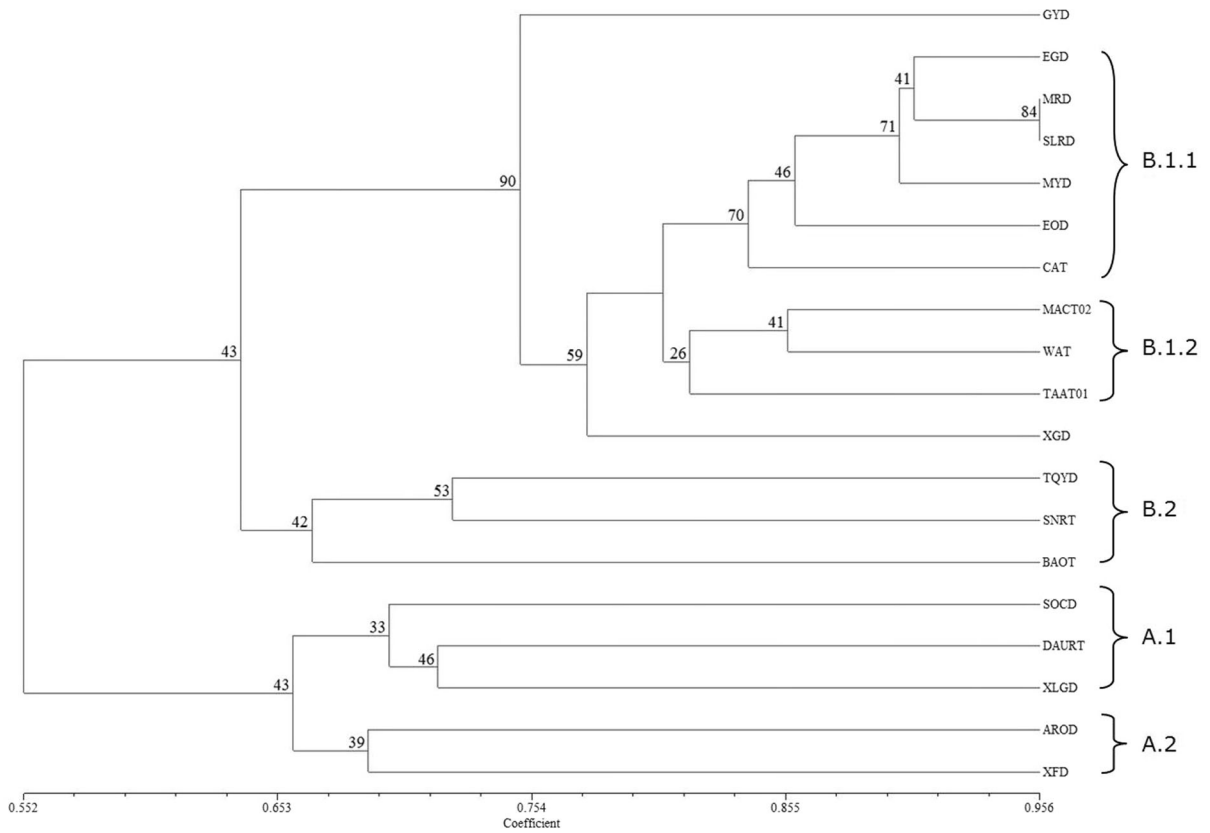


**Fig. 1** Three-dimensional representation of Principal Coordinate Analysis (PCoA) of pairwise Nei's genetic distances obtained by Genalex 6.51 software

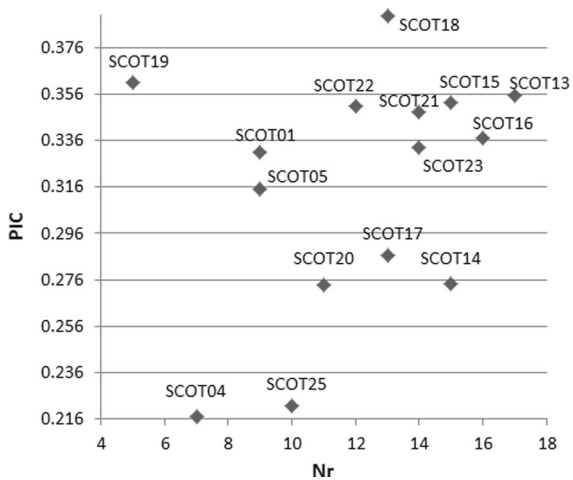
2009). This study suggested that SCoT primers can detect a high level of genetic diversity among coconut varieties, with an average PIC value of  $0.316 \pm 0.051$  comparable to that obtained by SSR and RAPD primers (Rivera et al. 1999; Meerow et al. 2003; Rajesh et al. 2014). Thus, in Fig. 3, by plotting a scatter graph of markers based on their PIC values and number of reproducible unique loci, we identified a group of markers that performed well on both measures (SCOT18, SCOT13, SCOT15, SCOT21, SCOT22 and SCOT23) in contrast to another group with low values for both measures (SCOT04 and SCOT25). The primers of the first group should therefore be prioritized in future coconut studies.

The allele distribution revealed by SCoT primer amplification indicated a high degree of diversity in

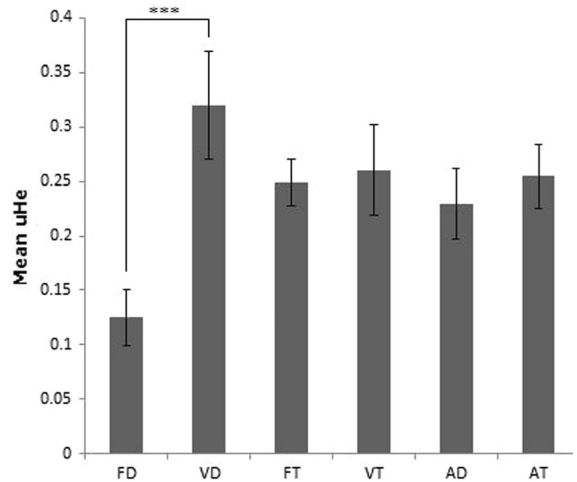
the seven Dwarf varieties originating from Vietnam (SOCD, EOD, TQYD, XGD, XFD, AROD and XLGD) compared to a relatively high degree of uniformity within each of the six non-native Dwarf varieties (GYD, EGD, CAT, MYD, MRD and SLRD), as shown in Fig. 4. The Nei's genetic diversity index (uHe) of Vietnamese Dwarfs was higher than the non-native Dwarf varieties. Applying the two-tailed t-test and assuming unequal variances, the Nei's genetic diversity index (uHe) of Vietnamese Dwarfs ( $M = 0.319$ ,  $SD = 0.066$ ) was higher than the non-native Dwarf varieties ( $M = 0.260$ ,  $SD = 0.086$ ),  $t(11) = 5.916$ ,  $p < 0.001$ . The Nei's genetic diversity index measures the probability that two individuals, chosen at random from the population, do not exhibit the same allele for the marker being studied.



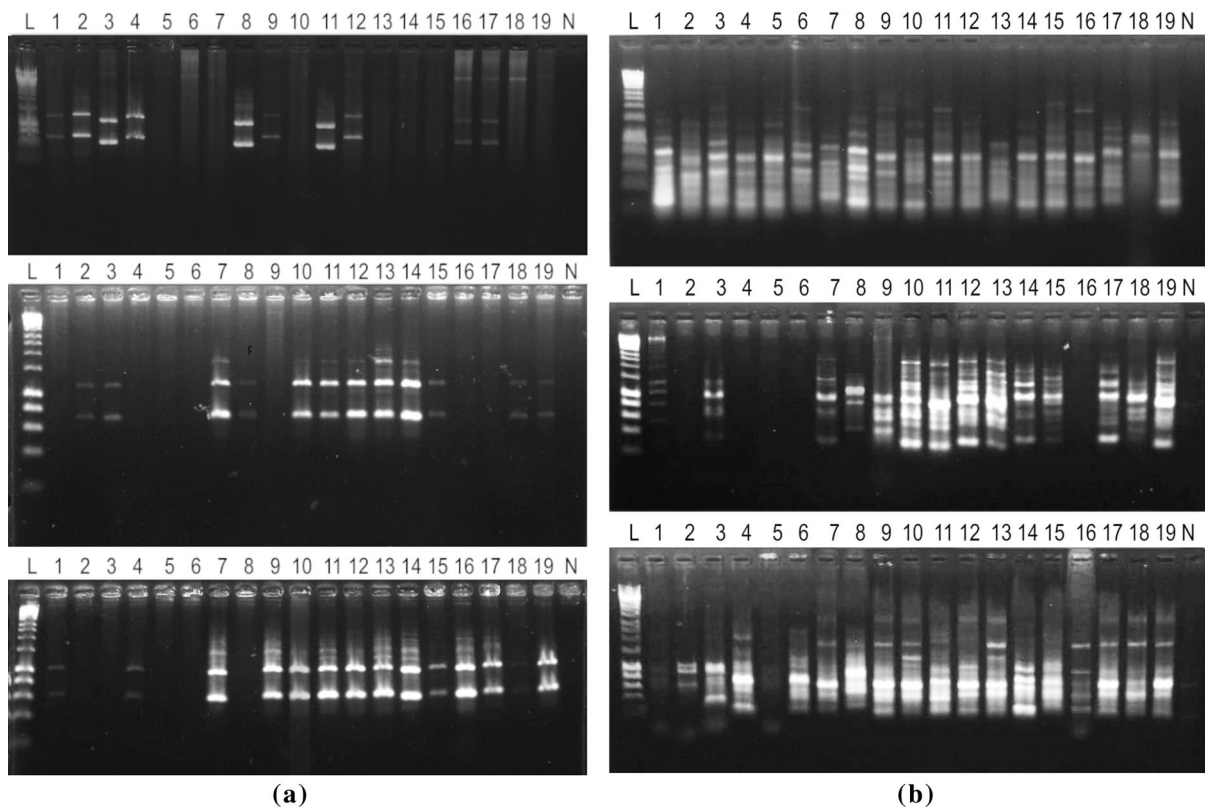
**Fig. 2** Dendrogram generated from the UPGMA clustering method using Simple Matching coefficients of 19 coconut varieties based on SCoT analysis. Numbers on the branches represent percentage bootstrap support based on 1000 iterations



**Fig. 3** Scatter plot of 15 SCoT primers based on values for PIC (Polymorphic Information Content) and Nr (Number of reproducible polymorphic bands)



**Fig. 4** Mean Nei's genetic diversity index (uHe) with standard error for foreign Dwarf (FD), Vietnamese Dwarf (VD), foreign Tall (FT), Vietnamese Tall (VT), all Dwarf (AD) and all Tall (AT) coconut varieties



**Fig. 5** Amplification profile of some SCOT primers in coconuts varieties used in this study **a** PCR amplification induced by SCOT04 primer, **b** PCR amplification induced by SCOT15 primer

Therefore, our results suggested that within-population diversity for Vietnamese Dwarf varieties was high (Fig. 5).

Previous genetic marker studies on coconut revealed that Dwarf varieties tend to be more homozygous by virtue of their self-pollinating nature, resulting in less variability (Perera et al. 1998, 2000). The sampling procedure for this study was prudently carried out in the most established germplasm collection in Vietnam. Thus, the high within-population diversity for Vietnamese Dwarf varieties revealed above suggests that traditional farming practices in Vietnam have greatly encouraged cross-fertilization. Unlike other regions where monoculture coconut plantations are well-established, most coconut palms in Ben Tre, Vietnam are planted as mixtures of varieties as small enclosures with high density, often intercropped with several fruits and vegetables. The high density of palms coupled with the great diversity of pollinating insects might have contributed to cross-

pollination even in traditionally self-pollinating Dwarfs. Dong Go Coconut Research Center selected native mother palms from surrounding small farms. Thus, results presented here most likely reflect the genetic diversity present in the Ben Tre region.

This high level of genetic diversity in Vietnamese Dwarf varieties has not been observed in previous studies. Compared to other regions such as China, America and Africa, Southeast Asia is regarded as the center of domestication of Dwarf coconuts. Hence, coconuts from this region tend to have a greater genetic diversity (Perera et al. 2000; Xiao et al. 2013). Nevertheless, in most studies that included Vietnamese Dwarfs, the genetic diversity of the Vietnamese group was not significantly different compared to populations from other Southeast Asian countries (Perera et al. 2003; Geethanjali et al. 2018). The allele distribution revealed by this study suggests that a closer examination of the genetic material of Vietnamese Dwarfs will be of interest, especially since

greater diversity allows for better and more sustainable resistance to diseases such as lethal yellowing (Baudouin et al. 2009).

Shannon's Information index was higher in Vietnamese Dwarf varieties ( $M = 0.471$ ,  $SD = 0.066$ ) than in non-native Dwarf varieties ( $M = 0.190$ ,  $SD = 0.052$ ),  $t(11) = 6.107$ ,  $p < 0.001$ . Shannon's index is a statistical measure of entropy, based on the weighted geometric mean of the frequency of each allele (Hennink and Zeven 1990; Spellerberg and Fedor 2003). If a marker has a greater number of alleles that are more evenly distributed across populations, this will result in a higher Shannon's index value. The lower number of alleles in non-native varieties might be due to the small number of qualified foreign mother palms conserved at Dong Go Coconut Research Center, leading to higher uniformity in these sample populations.

Among the exotic varieties, the Nei's genetic diversity index value for the Talls ( $M = 0.249$ ,  $SD = 0.072$ ) was not significantly higher than for the Dwarfs ( $M = 0.125$ ,  $SD = 0.051$ ),  $t(3) = 2.662$ ,  $p = 0.076$ . Previous studies suggested that the greater diversity in Tall varieties was due to allogamous breeding behaviors but the limited number of varieties in this study precludes a statistically significant conclusion (Rivera et al. 1999; Teulat et al. 2000). There was also no significant difference in Nei's index between the seven Vietnamese Dwarfs ( $M = 0.319$ ,  $SD = 0.066$ ) and the three Vietnamese Tall and hybrid varieties ( $M = 0.260$ ,  $SD = 0.086$ ),  $t(3) = 1.063$ ,  $p = 0.366$ , or between all Dwarf varieties ( $M = 0.230$ ,  $SD = 0.116$ ) and all Tall varieties ( $M = 0.255$ ,  $SD = 0.071$ ),  $t(15) = 0.575$ ,  $p = 0.574$ . This result further underscored the high degree of genetic diversity in Vietnamese Dwarf populations compared to previously studied populations.

Across all varieties, the moderate  $\Phi_{PT}$  value (9.41%) obtained from the AMOVA analysis indicated relatively little differentiation among coconut varieties.  $\Phi_{PT}$  is an analog of Wright's fixation index ( $F_{st}$ ). The value decreases with a greater degree of gene flow. A high value of  $\Phi_{PT}$  suggests significant barriers to inter-varietal breeding and, hence, high inter-varietal variance. By contrast, a lower value suggests the reverse, i.e. varieties are more similar to one another. Previous studies of coconut indicated that the degree of within-variety population was higher than among-variety, ranging from 53% in Sri Lankan

varieties (Perera et al. 1998), 60% in South Pacific varieties (Ashburner et al. 1997), 67% in Mexican varieties (Zizumbo-Villarreal et al. 2006) and up to 98% in Kenyan varieties (Oyoo and Muhammed 2016). The within-population variation partition of coconut varieties conserved in Dong Go Coconut Research Center was toward the higher end of the spectrum at 90.59%. This is primarily due to the greatly shared diversity among aforementioned Vietnamese Dwarf varieties, which in turn suggests a degree of introgression, especially among the Xiem family (XFD, XLGD, XGD and AROD).

Both the PCoA plot and UPGMA dendrogram delineated six non-native Dwarf varieties as a separate group. Both methods also indicated that the Vietnamese Dwarf varieties were similar to one another, and that the Tall varieties could be separated into distinct clusters (WAT, SNRT, BAOT, TAAT01 and MACT02). The Vietnamese hybrid DAURT was suggested by both methods to share more alleles with the Dwarf variety of XLGD. This result indicated that SCoT primers can provide useful information about the type and origin of coconut varieties, as suggested by the pioneering study on SCoT markers in coconut (Rajesh et al. 2015).

In conclusion, this study is one of the first to characterize Vietnamese coconut varieties using genetic markers. The high level of within-population genetic diversity revealed in Vietnamese coconuts, especially among the Dwarf varieties, suggested a higher degree of inter-breeding in Vietnamese Dwarfs compared to foreign Dwarf populations. This underscored the influence of traditional farming practice at small householder farms, where mixtures of coconut varieties are planted at high density among intercrops that attract pollinating insects. Future breeding efforts to improve yield and disease-resistance traits can benefit from further investigation of such diversity. SCoT markers with higher differentiation power were also identified as relatively able to demonstrate relationship among coconut varieties with Vietnamese origin or foreign origin, and likewise demonstrate relationship among Dwarfs and among Talls. The SCoT-PCR system thus proved useful in the exploratory assessment of diversity in large coconut germplasm collections.

**Acknowledgements** We are grateful to the Dong Go Coconut Research Center for generous provision of leaf materials and information.

**Author contributions** Conceptualization: NBQ, Data curation: NHMK, HNXM, NDNP, NBQ, Formal analysis: NHMK, Funding acquisition: NPT, Investigation: NHMK, NTQ, HNXM, NDNP, NPT and NBQ, Methodology: NBQ, NHMK, Validation: NBQ, NPT, Writing—original draft: NHMK, NBQ, Writing—review and editing: NBQ, NPT, NTQ.

**Funding** This research was funded by the Ministry of Science and Technology (MOST, Vietnam) under grant number ĐTĐL.CN-12/19.

**Data availability** Data supporting the findings of this study are available from the corresponding authors (Nguyen Bao Quoc and Nguyen Phuong Thao) upon reasonable request.

#### Declarations

**Conflict of interest** The authors of this article have no conflicts of interest.

**Consent for publication** All authors have given their consent for publication.

#### References

- Ashburner GR, Thompson WK, Halloran GM (1997) RAPD analysis of South Pacific coconut palm populations. *Crop Sci* 37(3):992–997
- Baudouin L, Philippe R, Quaiçoe R, Dery S, Dollet M (2009) General overview of genetic research and experimentation on coconut varieties tolerant/resistant to Lethal Yellowing. *Oléagineux Corps Gras Lipides* 16(2):127–131
- Chempakam B, Ratnambal MJ (1993) Variation for leaf polyphenols in coconut cultivars. In *Advances in Coconut Research and Development*. Oxford and IBH Publishing Co. Pvt. Ltd., New Delhi, India, pp. 51–53.
- Chen H, He XH, Luo C, Zhu J, Li F (2010) Analysis on the genetic diversity of 24 longan (*Dimocarpus longan*) accessions by SCoT markers. *Acta Horticulturae Sinica* 37:1651–1654
- Collard BC, Mackill DJ (2009) Start codon targeted (SCoT) polymorphism: a simple, novel DNA marker technique for generating gene-targeted markers in plants. *Plant Mol Biol Report* 27(1):86
- Dasanayaka PN, Nandadasa HG, Everard JMDT, Karunanayaka EH (2009) Analysis of coconut (*Cocos nucifera* L.) diversity using microsatellite markers with emphasis on management and utilisation of genetic resources. *J Nat Sci Found Sri Lanka* 37(2):99–109
- De Riek J, Calsyn E, Everaert I, Van Bockstaele E, De Loose M (2001) AFLP based alternatives for the assessment of distinctness, uniformity and stability of sugar beet varieties. *Theor Appl Genet* 103(8):1254–1265
- Duran Y, Rohde W, Kullaya A, Goikoetxea P, Ritter E (1997) Molecular analysis of East African tall coconut genotypes by DNA marker technology. *J Genet Breed* 51:279–288
- Everard JMDT, Katz M, Gregg K (1996) Inheritance of RAPD markers in the coconut palm *Cocos nucifera* L. *Trop Agric Res* 8:124–138
- 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 Gen Res* 16(2):156
- Gunn BF, Baudouin L, Olsen KM (2011) Independent origins of cultivated coconut (*Cocos nucifera* L.) in the Old World tropics. *PLOS ONE* 6(6):21143
- Guo X, Elston R (1999) Linkage information content of polymorphic genetic markers. *Hum Hered* 49(2):112–118
- Hayati PK, Hartana A (2000) Genetic diversity of Jombang coconut population based on RAPD markers. In *International Conference of Science and Technology for Managing Plant Genetic Diversity in the 21st Century*, pp. 12–16.
- Hennink S, Zeven AC (1990) The interpretation of Nei and Shannon-Weaver within population variation indices. *Euphytica* 51(3):235–240
- 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(1):103–108
- Lebrun P, Baudouin L, Bourdeix R, Konan JL, Barker JH, Aldam C, 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(6):962–970
- Lewontin RC (1972) The apportionment of human diversity. *Evol Biol* 6:381–398
- Loiola CM, Azevedo AON, Diniz LE, Aragao WM, Azevedo CDDO, Santos PHA, Ramos HCC, Pereira MG, Ramos SRR (2016) Genetic relationships among tall coconut palm (*Cocos nucifera* L.) accessions of the International Coconut Genebank for Latin America and the Caribbean (ICGLAC), evaluated using microsatellite markers (SSRs). *PLoS one* 11(3):0151309
- Manimekalai R, Nagarajan P, Kumaran PM (2006) Comparison of effectiveness of RAPD, ISSR and SSR markers for analysis of coconut (*Cocos nucifera* L.) germplasm accessions. *Trop Agric Res* 18:217–226
- Mauro-Herrera M, Meerow AW, Borrone JW, Kuhn DN, Schnell RJ (2007) Usefulness of WRKY gene-derived markers for assessing genetic population structure: an example with Florida coconut cultivars. *Sci Hortic* 115(1):19–26
- Meerow AW, Wissner RJ, Brown SJ, Kuhn DN, Schnell RJ, 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. *Theor Appl Gen* 106(4):715–726
- Meerow AW, Noblick L, Borrone JW, Couvreur TL, Mauro-Herrera M, Hahn WJ, Kun DN, Nakamura K, Oleas NH, Schnell RJ (2009) Phylogenetic analysis of seven WRKY genes across the palm subtribe *Attaleinae* (*Areaceae*) identifies *Syagrus* as sister group of the coconut. *PLoS One* 4(10):7353

- Nei M (1978) Estimation of average heterozygosity and genetic distance from a small number of individuals. *Genetics* 89(3):583–590
- Niral V, Jerard BA (2018) Botany, origin and genetic resources of coconut. In the coconut palm (*Cocos nucifera* L.)-research and development perspectives. Springer, Singapore
- Oyoo ME, Muhammed N, Cyrus KN, Githiri SM (2016) Assessment of the genetic diversity of Kenyan coconut germplasm using simple sequence repeat (SSR) markers. *Afr J Biotech* 15(40):2215–2223
- Peakall R, Smouse PE (2006) GENALEX 6: genetic analysis in excel. Population genetic software for teaching and research. *Mol Ecol Notes* 6:288–295
- Peakall R, Smouse PE (2012) GENALEX 6.5: genetic analysis in Excel. Population genetic software for teaching and research-an update. *Bioinformatics* 28:2537–2539
- Perera L, Russell JR, Provan J, McNicol JW, Powell W (1998) Evaluating genetic relationships between indigenous coconut (*Cocos nucifera* L.) accessions from Sri Lanka by means of AFLP profiling. *Theor Appl Gen* 96(3–4):545–550
- Perera L, Russell JR, Provan J, 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(1):15–21
- Perera L, Russell JR, Provan J, Powell W (2003) Studying genetic relationships among coconut varieties/populations using microsatellite markers. *Euphytica* 132(1):121–128
- 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
- Prevost A, Wilkinson MJA (1999) New system of comparing PCR primers applied to ISSR fingerprinting of potato cultivars. *Theor Appl Genet* 98(1):107–112
- Rajesh MK, Jerard BA, Preethi P, Thomas RJ, Karun A (2014) Application of RAPD markers in hybrid verification in coconut. *Crop Breed Appl Biotech* 14(1):36–41
- 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. *Biotech* 5(6):999–1006
- Rajesh MK, Sabana AA, Rachana KE, Rahman S, Ananda KS, Karun A (2016) Development of a SCoT-derived SCAR marker associated with tall-type palm trait in arecanut and its utilization in hybrid (dwarf x tall) authentication. *Ind J Gen Plant Breed* 76:119–122
- 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
- Rajesh MK, Karun A, Parthasarathy VA (2018) Coconut biotechnology. In the coconut palm (*Cocos nucifera* L.)-research and development perspectives. Springer, Singapore
- Reynolds KB, Cullerne DP, El Tahchy A, Rolland V, Blanchard CL, Wood CC, Singh SP, Petrie JR (2019) Identification of genes involved in lipid biosynthesis through de novo transcriptome assembly from *Cocos nucifera* developing endosperm. *Plant Cell Physiol* 60(5):945–960
- Rivera R, Edwards KJ, Barker JHA, Arnold GM, Ayad G, Hodgkin T, Karp A (1999) Isolation and characterization of polymorphic microsatellites in *Cocos nucifera* L. *Genome* 42:668–675
- Rohde W, Kullaya A, Rodriguez MJ, Ritter E (1996) Genome analysis of *Cocos nucifera* L. by PCR amplification of spacer sequences separating a subset of Copia-like Ecorireplicative elements. *Philippine J Crop Sci (philippines)* 21:26–31
- Rohde W, Sniady V, Herran A, Estioko L, Sinje S, Marseillac N, Angeliq B, Patricia L, Dieter P, Alois K, Judith R, Norbert B, Rodriguez J (2002) Construction and exploitation of high density DNA marker and physical maps in the perennial tropical oil crops coconut and oil palm: from biotechnology towards marker-assisted breeding. *Burotrop Bulletin* 20:13–14
- Rohlf FJ (2002) NTSYSpc: Numerical taxonomy system, ver. 2.1. Setauket, NY: Exeter Publishing Ltd.
- Saensuk C, Wanchana S, Choowongkomon K, Wongpornchai S, Kraithong T, Imsabai W, Chaichoompu E, Ruanjaichon V, Toojinda T, Vanavichit A, Arikrit S (2016) De novo transcriptome assembly and identification of the gene conferring a “pandan-like” aroma in coconut (*Cocos nucifera* L.). *Plant Sci* 252:324–334
- Shalini KV, Manjunatha S, Lebrun P, Berger A, Baudouin L, Pirany N, Raganath RM, Prasad DT (2007) Identification of molecular markers associated with mite resistance in coconut (*Cocos nucifera* L.). *Genome* 50(1):35–42
- Spellerberg IF, Fedor PJA (2003) A tribute to Claude Shannon (1916–2001) and a plea for more rigorous use of species richness, species diversity and the ‘Shannon–Wiener’ Index. *Glob Ecol Biogeogr* 12(3):177–179
- 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
- Upadhyaya 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
- Xiao Y, Luo Y, Yang Y, Fan H, Xia W, Annaliese S, Mason AS, Zhao S, Sager R, 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* 6(3):193–200
- Yang Y, Bocs S, Fan H, Armero A, Baudouin L, Xu P, Xu J, This D, Hamelin C, Iqbal A, Qadri R, Zhou L, Li Z, Wu Y, Ma Z, Issali AE, Rivallan R, Liu N, Xia W, Peng M, Xiao Y (2021) Coconut genome assembly enables evolutionary analysis of palms and highlights signaling pathways involved in salt tolerance. *Commun Biol* 4(1):1–14
- Yeh FC, Yang RC, Boyle TB, Ye ZH, Mao JX (1997) Popgene, version 1.32; the user-friendly shareware for population genetic analysis. *Can Univ Alberta Mol Biol Biotechnol Centre* 10:295–301
- Zhang J, Guo D, Gong Y, Liu C, Li M, Zhang G (2011) Optimization of start codon targeted polymorphism PCR

(SCoT-PCR) system in *Vitis vinifera*. J Fruit Sci 28(2):209–214

Zizumbo-Villarreal D, Ruiz-Rodriguez M, Harries H, Colunga-GarciaMarin P (2006) Population genetics, lethal yellowing disease, and relationships among mexican and imported coconut ecotypes. Crop Sci 46:2509–2516

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.