Tree Genetics & Genomes (2016) 12:116
DOI 10.1007/s11295-016-1076-x

ORIGINAL ARTICLE

SSR markers reveal the population structure of Sri Lankan yellow dwarf coconuts (*Cocos nucifera* L.)

L. C. J. Kamaral[1] · P. N. Dassanayaka[2] · K. L. N. S. Perera[3] · S. A. C. N. Perera[1]

Received: 30 October 2015 / Revised: 15 September 2016 / Accepted: 7 November 2016
© Springer-Verlag Berlin Heidelberg 2016

Abstract World coconut germplasm has been classified broadly as tall and dwarf coconuts based on palm stature. Dwarf coconuts are predominantly self-breeding purelines hypothesised to have derived from tall coconuts. Dwarfs are categorized as yellow, green, red and brown on the colour of epicarp and are important as parents in hybridization of coconuts for desirable traits. Sri Lankan yellow dwarfs (SLYD) were observed to have uncommon phenotypes which were not previously reported for dwarf coconuts in the world, and this study was conducted to elucidate the population structure of SLYD. One hundred and two randomly selected SLYD individuals were categorized into three morphological groups and their genotypes were derived at 30 SSR loci. Genotypic data were analysed in PowerMarker 3.2.5 and Structure 2.3.4 software to derive the genetic diversity and the population structure. Unexpectedly high numbrs of alleles, genotypes, gene diversity and heterozygosity values were recorded for SLYD. Four populations were identified within SLYD under admixture model and their morphological variations were determined. Cross pollination between the dwarf and tall coconut varieties followed by the fixing of alleles by subsequent self-pollination was hypothesised to be the cause for the emergence of new genetic groups within dwarf populations. The study demonstrated the formation of new genotypes upon limited cross pollination of even naturally self-pollinating tree crops. The information will be useful for developing strategies for germplasm conservation, practical coconut breeding and determining the domestication and evolution of dwarf coconuts.

Keywords Dwarf coconuts · Evolution · Genetic diversity · Self-pollination · Cross pollination ·

Communicated by W. Ratnam

✉ S. A. C. N. Perera
chandrikaperera2003@yahoo.com

[1] Coconut Research Institute of Sri Lanka, Bandirippuwa Estate, Lunuwila 61150, Sri Lanka

[2] University of Sri Jayawardenepura, Gangodawila, Nugegoda, Sri Lanka

[3] Genetech Molecular Diagnostics, No.54, Kithulwatta Rd, Colombo 08, Sri Lanka

Introduction

The coconut palm, *Cocos nucifera* L., is a member of the family Arecaceae (palm family) and is the sole species in the genus *Cocos*. It is an extensively grown tropical plantation and small holder tree crop offering a multitude of uses, oil being the longest-term product. However, at present, the interest on coconut palm has expanded well beyond the tropics for its diverse uses mainly as a healthy natural beverage and for natural coir and fibre products in addition to its traditional uses.

Two types of coconuts that differ in morphology and in breeding behaviour are found in the world. The tall coconuts are tall in habit and display cross pollinating breeding behaviour resulting in allogamous individuals and heterogenous populations. The tall coconuts are late flowering, more tolerant to harsh environmental conditions and produce larger nuts. Tall coconuts are the most favoured in the plantation scale cultivation around the world. The dwarf coconuts are shorter and are predominantly self-pollinating producing homozygous purelines. The dwarf coconuts are more precocious, susceptible to harsh environments and produce smaller nuts but in large numbers. Dwarf coconuts are the predominant choice among small holder coconut growers as well as in home gardens in tropical landscape and also the most preferred coconut variety

Published online: 14 November 2016

for natural beverage purposes. The hybrids between these two coconut types are the most favoured among the coconut growers for their combined, economically more viable traits.

Southeast Asia and South Asia including Sri Lanka are hypothesised to be centres of origin for tall coconuts (Gunn et al. 2011). The dwarf coconuts are presumed to be evolved from the tall coconuts of the Southeast Asia with a reduction in allelic richness (Perera et al. 2003). Due to self-pollination, fruit colour in dwarf cultivars is uniform and they are referred to as either green, yellow, red or brown. Out of them, yellow dwarf and red dwarf have been extensively used in the production of popular commercial hybrids such as MAYPAN, MATAG and CRIC65, in several countries. Malayan Yellow is the most widely known yellow dwarf coconut population in the world. In addition, yellow dwarf coconuts have been reported in several other countries such as Sri Lanka, India and the Philippines. Most of these yellow dwarf populations conform to the typical dwarf morphological characters, listed as short stature, smaller crown, drooping fronds, lack of a root bole (which is the swollen base of the trunk), etc. However, the observations of the Sri Lanka yellow dwarf coconut germplasm repository proved not to conform to the characteristic morphological features suggesting a population structure and raising hypothesis on a different origin or domestication of yellow dwarf coconuts within the country. The aim of the present investigation was to determine the population structure of the yellow dwarf coconut population conserved ex situ in the field gene bank of coconut in Sri Lanka to enhance the knowledge on the genetics of yellow dwarf coconuts in the world for refining conservation practices and for more efficient applications in the genetic improvement of this tree crop.

Materials and methods

Sample palms

A total of 102 SLYD palms which are planted at the field gene bank at Ambakelle, Sri Lanka, were randomly selected for this study. This particular germplasm repository is the second generation of open pollinated progeny of yellow dwarfs that have been conserved ex situ. The original ex situ conserved progeny was derived from open pollinated in situ collections spanning throughout the country to adequately represent the yellow dwarf coconut population found within Sri Lanka. In addition, three each of Sri Lanka tall (SLT), Gon thembili tall (GTT-Sri Lankan indigenous tall coconut form bearing yellow coloured nuts) and Sri Lanka green dwarf (SLGD) were sampled for comparison purposes. In the analysis sample, reference numbers 1 to 102 refer to SLYD sample palms while 103–105, 106–108 and 109–111 refer to SLT, SLGD and GTT sample palms, respectively.

The morphology of the 102 sample palms was observed and they were categorized into three morphological groups based on the appearance of the crown and the absence/presence of a root bole. Palms with larger crown and non-drooping rachis ends and having root boles were included in tall-like (TL) group, palms with smaller crown and drooping rachis ends and lacking root boles were included in dwarf-like (DL) group, and palms with mixed characters were included in the mixed (M) group. The pollination behaviour was observed in the selected individuals of the SLYD and 10 palms of the comparative Sri Lanka green dwarf coconut variety and the durations of the male and female phases were also recorded.

DNA extraction

Leaflet segments of approximately 0.5 g were taken from the immature leaf tissues of each sampled tree. The DNA was extracted following a CTAB protocol developed by Weising and Kahl (1997) and modified by Doyle and Doyle (1990).

SSR analysis

A total of 30 microsatellite primer pairs were used for genotyping. These primers consisted of ten CAC markers (Perera et al. 2003), ten CNZ markers (Rivera et al. 1999) and ten CnCir markers (Baudouin and Lebrun 2002). Primer sequences, sizes of PCR products and annealing temperatures for each primer pair are given in Table 1.

PCR reaction was performed in a final volume of 25 μL which contained a mixture of 4 μL of (20 ng/μL) template DNA, 1 × *Taq* PCR green buffer containing 2 mM $MgCl_2$, 1.25 U of *Taq* DNA polymerase (Dream *Taq*-Fermentas), deoxynucleoside triphosphates (0.35 mM each; Geneshun Biotech) and 0.6-μM primer pair (1st BASE). PCR reaction cycles consisted of 4-min initial denaturation at 94 °C followed by 35 cycles of 30 s at 94 °C, 30 s at annealing temperature (depending on primer) and 1 min at 72 °C, and final extension at 72 °C for 5 min. PCR was performed using thermal cycler (Applied Biosystems).

PCR product was subjected to 6% denaturing polyacrylamide gel electrophoresis followed by silver staining (Anolles and Petter 1994). The genotypes were obtained by allele size (number of base pairs) differences in comparison to a standard marker (50 bp).

Statistical analysis

PowerMarker version 3.25 (Liu and Muse 2005) was used to calculate summary statistics number of genotypes, number of alleles per locus, heterozygosity and gene diversity (probability that two randomly selected alleles from the population are different/expected heterozygosity) which were used to calculate the inbreeding coefficient and to measure the within

Table 1 Details of microsatellite primers used for genotyping of coconut palms

Oligo name		Sequence	Size (bp)	Annealing temperature (°C)
CAC02	F	5′-AGC TTT TTC ATT GCT GGA AT-3′	210–254	58
	R	5′-CCC CTC CAA TAC ATT TTT CC-3′		
CAC03	F	5′-GGC TCT CCA GCA GAG GCT TAC-3′	187–203	55
	R	5′-GGG ACA CCA GAA AAA GCC-3′		
CAC04	F	5′-CCC CTA TAG ATC AAA ACA AG-3′	182–216	58
	R	5′-CTC AGT GTC CGT CTT TGT CC-3′		
CAC06	F	5′-TGT ACA TGT TTT TTG CCC AA-3′	150–168	52
	R	5′-CGA TGT AGC TAC CTT CCC C-3′		
CAC08	F	5′-ATC ACC CCA ATA CAA GGA CA-3′	188–210	56
	R	5′-AAT TCT ATG GTC CAC CCA CA-3′		
CAC10	F	5′-GGA ACC TCT TTT GGG TCA TT-3′	195–205	56
	R	5′-GAT GGA AGG TAA TGC TG-3′		
CAC20	F	5′-CTC ATG AAC CAA ACG TTA GA-3′	124–133	54
	R	5′-CAT CAT ATA CAT ACA TGC AAC A-3′		
CAC21	F	5′-AAT TGT GTG ACA CGT AGC C-3′	149–151	54
	R	5′-GCA TAA CTC TTT CAT AAG GGA3′		
CAC23	F	5′-TGA AAA CAA AAG ATA GAT GTC AG-3′	170–179	56
	R	5′-GAA GAT GCT TTG ATA TGG AAC-3′		
CAC65	F	5′-GAA AAG GAT GTA ATA AGC TGG-3′	150–173	54
	R	5′-TTT GTC CCC AAA TAT AGG TAG-3′		
CNZ04	F	5′-TAT ATG GGA TGC TTT AGT GGA-3′	130–166	53
	R	5′-CAA ATC GAC AGA CAT CCT AAA-3′		
CNZ06	F	5′-ATA CTC ATC ATC ATA CGA CGC-3′	69–97	52
	R	5′-CTC CCA CAA AAT CAT GTT ATT-3′		
CNZ10	F	5′-CCT ATT GCA CCT AAG CAA TTA-3′	108–152	56
	R	5′-AAT GAT TTT CGA AGA GAG GTC-3′		
CNZ12	F	5′-TAG CTT CCT GAG ATA AGA TGC-3′	218–229	54
	R	5′-GAT CAT GGA ACG AAA ACA TTA-3′		
CNZ21	F	5′-ATG TTT TAG CTT CAC CAT GAA-3′	220–250	54
	R	5′-TCA AGT TCA AGT TCA AGA AGA CCT TTG-3′		
CNZ29	F	5′-TAA ATG GGT AAG TGT TTG TGC-3′	105–157	56
	R	5′-CTG TCC TAT TTC CCT TTC ATT-3′		
CNZ40	F	5′-CTT GAT TGC TAT CTC AAA TGG-3′	143–155	56
	R	5′-CTG AGA CCA AAT ACC ATG TGT-3′		
CNZ43	F	5′-TCT TCA TTT GAT GAG AAT GCT-3′	175–219	54
	R	5′-ACC GTA TTC AAC ATT CTA ACA-3′		
CNZ44	F	5′-CAT CAG TTC CAC TCT CAT TTC-3′	151–170	52
	R	5′-CAA CAA AAG ACA TAG GTG GTC-3′		
CNZ46	F	5′-TTG GTT AGT ATA GCC ATG CAT-3′	101–120	56
	R	5′-AAC CAT TTG TAG TAT ACC CCC-3′		
CnCir01	F	5′-TTG GTC TAT TGC ATG TTC-3′	150	44
	R	5′-TGG CAT TGA GAG GGT-3′		
CnCirC5	F	5′-ACC ACC AAA GCC AGA GC-3′	133	50
	R	5′-GCA GCC ACT ACC TAA AAA G-3′		
CnCirE4	F	5′-GCA TGG TAT TCG GAT TTG-3′	200	50
	R	5′-ATG GTT CAG ATT TGG ACA GT-3′		
CnCirA4	F	5′-GTT GGT TAC TGG AAA TCT T-3′	196–204	50
	R	5′-CAT GAC ATA CGG ACT AGC-3′		
CnCirB4	F	5′-TTT CAT TCC AAG AGC CTA C-3′	200	50
	R	5′-TTT CAA GCA TCA TTT CAA CT-3′		
CnCirD8	F	5′GCT CTT GAT GTG GCT GCT-3′	250	54
	R	5′-AGG CGT GTT GAG ATT GTG A-3′		
CnCirB11	F	5′-TCT GCA TCC CTT CTT TAT TA-3′	225	52
	R	5′-TTG TCT TTC TTT ATT CTA TTG G-3′		

Table 1 (continued)

Oligo name		Sequence	Size (bp)	Annealing temperature (°C)
CnCirHll	F	5′-TCA TTC AGA GGA CAA AAG TT-3′	150–200	46
	R	5′-TAA AAA TTC ATA AAG GTA AAA-3′		
CnCir51	F	5′-TCT CGT GGA TCT CGT C-3′	200	48
	R	5′-GCT CTT CCA GTT ACG TTT-3′		
CnCir89	F	5′-GAG TTG GAG AAG AAG AGG-3′	350	48
	R	5′-ACG ACA ATA GAT GGA ACA-3′		

population genetic diversity of SLYD coconut population. Shared allele-based genetic distances were calculated among the individuals including the three comparative genotypes and UPGMA (unweighted pair group method with arithmetic mean) dendrogram was constructed using genetic distances among individuals derived by Neil's method (Nei et al. 1983) in PowerMarker software.

Population clustering was performed using multi-locus genotypic data in the software program STRUCTURE version 2.3.4 (Pritchard et al. 2000). This software assumes population structure and uses interactive algorithm to allocate individuals (probabilistically) into K clusters, based on clustering method and assumes a model in which there are K populations. The data were analysed under admixture model and each run was done for $K = 1$ to $K = 12$, 20,000 burn period and 20,000 MCMC repeats after burn. Graphical representations of STRUCTURE results were produced using the programme DISTRUCT (Rosenberg 2004).

Results

Morphological categorization

Out of the 102 randomly selected sample palms of the SLYD population, 30, 42 and 30 were identified in tall-like, dwarf-like and mixed character groups, respectively, based on the stem and crown morphology as described in the methodology. The study on the pollination behaviour revealed all the SLYD as well as observed green dwarf palms to be self-pollinating as a result of the overlapping of male and female phases within the same inflorescence. Accordingly, the self-pollinating breeding behaviour which is the characteristic of dwarf coconuts was confirmed in the SLYD population also despite their morphological variations. Similarly, the studied SLGD also recorded the self-pollinating breeding behaviour while the naturally cross pollinating breeding behaviour in tall coconuts has long been established.

Summary statistics of genotypic data

Out of a total of 30 SSR loci, 29 were revealed to be polymorphic for the SLYD population while the marker locus CnCir89 was observed to be monomorphic. A total of 100 alleles were scored within SLYD, with an average of 3.37 alleles per locus, forming 148 genotypes with marker locus CAC06 being the most informative recording the highest number of alleles (06) (Table 2). Marker loci CAC20, CAC21, CAC23, CNZ10, CNZ29, CnCirE4 and CnCirD8 scored only two alleles at each locus. The range of summary statistics at certain SSR loci (numbers of alleles and genotypes, gene diversity and expected heterozygosity) was relatively high which is not expected in a self-breeding pureline population (Table 2).

Mean gene diversitiy and heterozygosity values of the three morphological groups, TL, DL and M, were calculated and the inbreeding coefficients (F_{IS}) were derived for the same (Table 3). The highest F_{IS} value was observed in DL group confirming high homozygosity within the DL group following the general situation in dwarf coconuts. The highest F_{IS} was recorded in TL group with the M group recording a closer value to that of TL indicating relatively high levels of heterozygosity in these morphological groups.

Genetic distances and the phylogenetic relationships

The dendrogram drawn for the sample SLYD individuals and the three control varieties revealed different clusters within SLYD (Fig. 1). The dendrogram is subdivided into 4 clusters for easy reference in the analysis. Accordingly, the main subcluster (cluster 1) was located (palm numbers 104 to 38 clockwise) in the right side of the dendrogram. To the right of cluster 1, there are 18 yellow dwarf palms (cluster 2) including palm numbers 63–19 clockwise. Cluster 3, located in clockwise direction, includes 26 palms (numbers 61 to 49) while cluster 4 includes 57 individuals representing palm numbers 29 to 2 in clockwise direction. Cluster 1 in the dendrogram comprises four yellow dwarfs (37, 38, 53 and 54) and tall coconuts, SLT (103–105) and GTT (109–111) indicating allele sharing of yellow dwarf with tall coconuts in Sri Lanka. Cluster 2 included 12, 2 and 4 SLYD individuals representing TL, M and DL morphological groups, respectively. The three individuals of the control variety SLGD (numbers 106–108) were included in cluster 3 along with 15 SLYD individuals (representing the DL morphological group (except for 61, 62, 97, 80, 79 which belonged to M morphological group). The leftmost part of the dendrogram (numbers 29–2 clockwise)

Table 2 Summary statistics of genotypic data of SLYD

Marker	No. of genotypes	No. of alleles	Gene diversity	Heterozygosity	F_{IS}
CAC02	7	4	0.2029	0.0659	0.675
CAC03	3	3	0.0586	0.0600	−0.023
CAC04	11	5	0.5630	0.0968	0.828
CAC06	6	6	0.1992	0.1327	0.334
CAC08	6	4	0.2275	0.0900	0.604
CAC10	4	3	0.1482	0.0737	0.503
CAC20	2	2	0.0605	0.0625	−0.033
CAC21	2	2	0.0200	0.0202	−0.01
CAC23	3	2	0.1556	0.0700	0.550
CAC65	11	5	0.5732	0.1134	0.802
CNZ04	6	4	0.3151	0.0722	0.771
CNZ06	8	4	0.5222	0.0612	0.883
CNZ10	2	2	0.4588	0.0000	1.0
CNZ12	4	3	0.2164	0.1146	0.470
CNZ21	5	3	0.4183	0.0928	0.778
CNZ29	2	2	0.0204	0.0206	−0.01
CNZ40	4	3	0.2423	0.0714	0.705
CNZ44	6	4	0.3854	0.1064	0.724
CNZ43	5	3	0.2613	0.1400	0.464
CNZ46	4	3	0.2494	0.0300	0.880
CnCir01	7	5	0.2358	0.0309	0.869
CnCir51	6	4	0.2732	0.0909	0.667
CnCirE4	3	2	0.2169	0.0495	0.772
CnCirC5	6	5	0.5438	0.1000	0.816
CnCirB4	4	3	0.2463	0.1546	0.372
CnCirA4	6	4	0.3074	0.0825	0.732
CnCirH11	8	5	0.2835	0.1735	0.388
CnCirD8	3	2	0.4550	0.0200	0.956
CnCirB11	4	3	0.1392	0.1053	0.243
Total	148	100	–	–	–
Mean	4.9667	3.3667	0.2666	0.0767	

F_{IS} inbreeding coefficient

consisted of a large number of small clusters of palms representing the three groups DL, TL and mixed morphology indicating the complex genetic composition of the studied population.

Population structure analysis

Structure analysis provided evidence for the presence of a population structure within SLYD population. Results were obtained separately for $K = 1$ to $K = 12$ populations for the individuals of SLYD population and the same K levels including the three reference coconut forms also.

Including the two reference tall coconut forms, the optimum population number was recorded to be ($K = 5$) under admixture model (Fig. 2). Structure analysis divided the SLYD population into 4 clusters while the SLT and GTT formed a separate cluster (named ST 1 to ST 5). Allele frequency divergence among populations computed using point estimates of P (Nei's minimum distance) and the average distances between individuals of the same cluster and F_{ST} (effect of subpopulations compared to the total population) derived by the Structure software are given in Tables 4 and 5, respectively.

Table 3 Summary statistics of different morphological groups within SLYD population

Morphological group	Mean gene diversity	Mean heterozygosity	F_{IS}
DL	0.1956	0.0376	0.808
TL	0.3093	0.1107	0.642
M	0.2577	0.0933	0.638

F_{IS} inbreeding coefficient

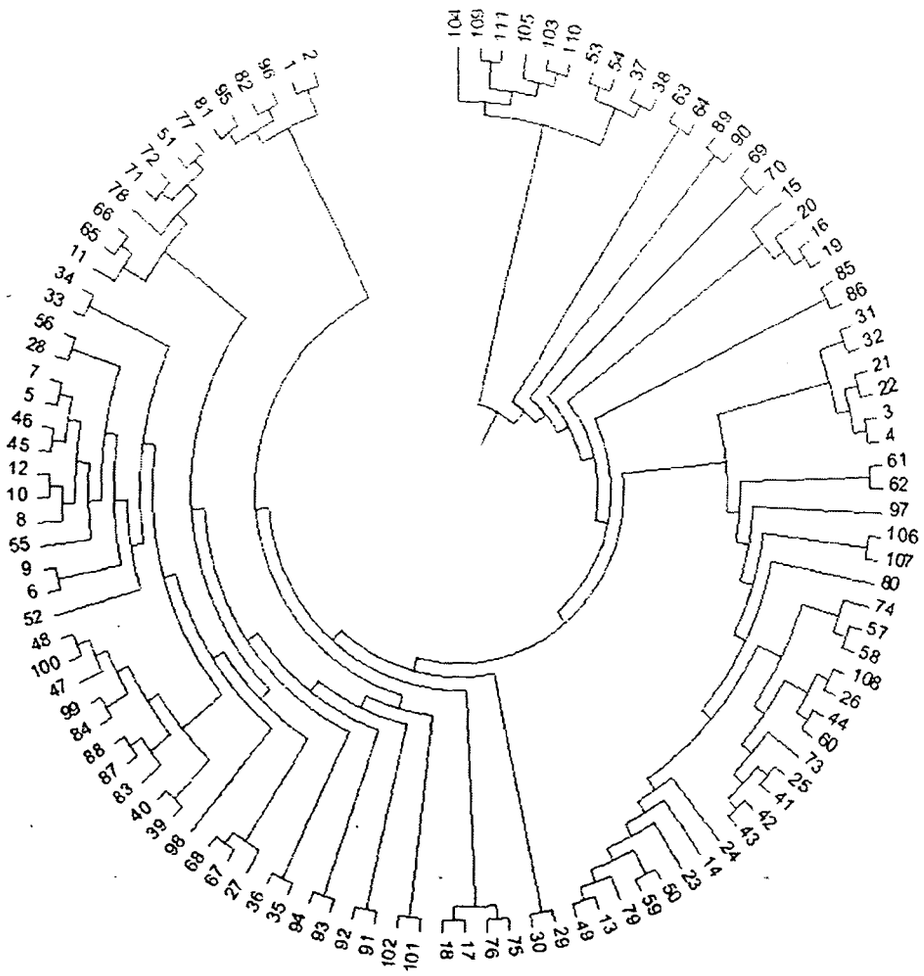

Fig. 1 UPGMA phylogenetic tree drawn based on shared allele distances. Sample numbers 103, 104 and 105 are SLTI; samples 109, 110 and 111 are GTT; and samples 106,107 and 108 are SLGD

As per Table 4, the highest allele frequency divergence was observed between ST 3 and ST 5. ST 3 comprised of the control SLT, control GTT and 4 sample palms of SLYD while ST 5 comprised of 50 SLYD individuals. ST 1 included the control SLGD along with 18 and 4 SLYD individuals categorized in DL and M morphological groups, respectively. Consequently, the lowest expected heterozygosity was recorded in ST 1 while the same cluster recorded the highest F_{ST} value (Table 5) indicating fixation of alleles due to high inbreeding in this subpopulation.

The morphological groups, DL, TL and M, were compared with the different clusters resulted from structure analysis (ST 1 to ST 5) as well as the dendrogram (Clusters 1 to 4) derived with Powermarker (Table 6).

The morphological, structural and diverse groups in the dendrogram reported comparable results although not matching exactly. Accordingly, cluster 1 and ST 3 were comparable including the SLT, GTT and 4 SLYD individuals belonging to DL and M groups with exact matching. Cluster 2 included mostly the TL group and was comparable with ST 4 in structure analysis. Nearly 50% of the DL individuals belonged to dengrogram cluster 3 and this was matching with ST 1 in the structure analysis. This cluster included control SLGD also and helped in distinguishing the pure yellow dwarf individuals from the mixed population. Cluster 4 in the dendrogram was the largest including 57 individuals belonging to all the three morphological groups. Individual palms in this cluster were comparable with the largest structure group ST 5 while it shared the SLYD individuals of ST 2 also.

Discussion and conclusions

Within population diversity

Dwarf coconuts in the world are known to be self-breeding homozygous purelines. Results of the current study revealed a high number of alleles, genotypes, gene diversity and heterozygosity values which are not expected in a dwarf population providing molecular evidence for the morphological non-uniformity of the SLYD population.

In the present study, an average of 3.4 alleles per SSR locus was scored. This value is high compared with the previous studies done for global as well as Sri Lankan coconut

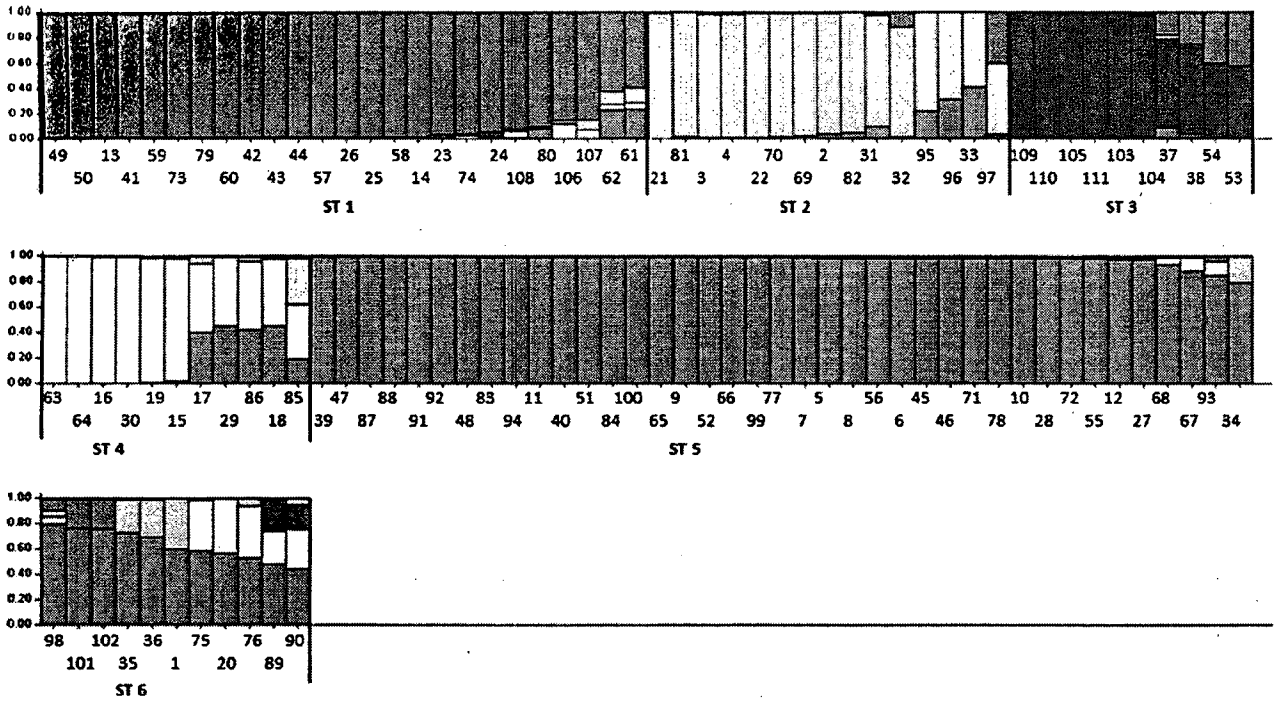

Fig. 2 Population structure of SLYD and the reference coconut forms. *DL* dwarf-like morphology, *TL* tall-like morphology, *SLT* reference tall varieties, *MIXED* mixed DL and TL morphology)

germplasm. Meerow et al. (2003) in a study using 15 SSR markers reported average allele numbers of 2.8 for Atlantic tall, 3.0 for Panama tall, 1.33 for Malayan red dwarf, 1.6 for Malayan yellow dwarf, 1.8 for Malayan green dwarf and 2.73 for Fiji dwarf per SSR marker locus. Dasanayaka et al. (2009) analysed Sri Lankan coconut germplasm using 16 SSR markers and reported average allele frequencies of 4.75 for tall, 1.81 for dwarf and 1.87 for king coconut (Sri Lanka intermediate coconut) per marker locus. Accordingly, in comparison with the results obtained by Meerow et al. (2003) and Dasanayaka et al. (2009), it is confirmed that the allele richness obtained in the current study for SLYD is much higher than all the dwarfs, and even higher than some of the tall varieties studied in the above studies.

Gene diversity has been observed to be higher in tall coconut populations than in dwarf populations (Dasanayaka et al. 2009; Perera et al. 2003). Perera et al. (2003) in a study representing world coconut populations reported an overall mean gene diversity values of 0.703 and 0.37 for tall and dwarf, respectively. Dasanayaka et al. (2009) studied Sri Lankan coconut germplasm and recorded an overall mean gene diversity of 0.55 and 0.21 for tall and dwarf coconuts, respectively. The observed gene diversity of SLYD (0.27) is higher compared with the results of the above studies especially considering that only one particular dwarf (SLYD) has been analysed in the present study in contrast to higher numbers of dwarf coconuts studied by Perera et al. (2003) and Dasanayaka et al. (2009). Observed heterozygosity which ranged from 0.0200 to 1.000, although was higher than expected from a self-breeding pureline, was low compared with the gene diversity indicating rare alleles at several SSR loci.

Several different clusters were observed in the UPGMA dendrogram indicating a genetic structure within the SLYD population. Perera et al. (2000) included two SLYD palms in a study to compare the world coconut germplasm and these two palms separated into two clusters. One of them belonged in a group that included most of the dwarf varieties, especially Sri Lanka intermediate (king coconut), Sri Lanka brown

Table 4 Allele frequency divergence among clusters

	ST 1	ST 2	ST 3	ST 4	ST 5
ST 1	–	0.1397	0.5407	0.2693	0.1745
ST 2	0.1397	–	0.4519	0.1943	0.0868
ST 3	0.5407	0.4519	–	0.4405	0.5451
ST 4	0.2693	0.1943	0.4405	–	0.2348
ST 5	0.1745	0.0868	0.5451	0.2348	–

Table 5 Expected heterozygosity (average distance between individuals of the same cluster) and F_{ST} of clusters

Cluster	Expected heterozygosity	F_{ST}
ST 1	0.0497	0.8380
ST 2	0.2583	0.4898
ST 3	0.2876	0.6631
ST 4	0.2715	0.5376
ST 5	0.1079	0.7476

Table 6 Comparison of morphological groups (DL, TL and M) with clusters (1 to 4) formed in the dendrogram and clusters derived in Structure (ST 1 to ST 4)

MG	TN	Clusters in dendrogram				Structure Clusters (ST 1 to ST 5)				
		1	2	3	4	ST 1	ST 2	ST 3	ST 4	ST 5
DL	42	2	4	18	18	18	3	2	1	18
TL	30	0	12	0	18	0	6	0	8	16
M	30	2	2	5	21	4	6	2	2	16
SLT	3	3	0	0	0	0	0	3	0	0
GTT	3	3	0	0	0	0	0	3	0	0
SLGD	3	0	0	3	0	3	0	0	0	0
	111	10	18	26	57	25	15	10	11	50

MG morphological group, *TN* total number

dwarf, Thailand green dwarf, Thailand red dwarf, Indonesian green dwarf, African yellow dwarf, etc., while the other grouped with most of the tall varieties including Indonesian tall, Pacific tall, Vietnam tall and Thailand tall giving clues for the population variation of SLYD. However, the sample sizes of the previous studies were quite inadequate to confirm the findings.

Population structure

Structure analysis performed in the current study revealed the genetic structure of SLYD population separating out four groups within population under admixture model. Morphologically, the four groups consisted of a pure dwarf-like group (ST 1), a pure tall-like group (ST 4) and two mixed groups (ST 2 and ST 5) possessing mixed morphologies of tall and dwarf coconuts. Variations between populations have been observed to be higher than within population variation in inbreeding varieties (Perera et al. 1998). But SLYD although is inbreeding as confirmed in the current study shows high variation and a genetic structure within population. High levels of admixture have previously been commonly observed in tall coconut forms and have not been detected in dwarf coconut forms (Perera et al. 2003). However, it is shown that the SLYD includes genetically admixed individuals with divergent phenotypes, which is contradictory when compared with other dwarf coconut forms in Sri Lanka and in the world.

Dwarf coconuts in the world have been originated from the tall coconuts in the Southeast Asian and Pacific region (Perera et al. 2003). Thus, the Sri Lankan dwarf coconuts have not evolved from the Sri Lankan tall coconuts but represent a separate introduction (Perera et al. 2000) and they share alleles with the tall coconuts from the Southeast Asian and the Pacific region rather than the tall coconuts from South India (India and Sri Lanka). No records are available on the time periods or the places of introduction of dwarf coconuts into Sri Lanka.

A small percentage of outcrossing generally occurs in any self-pollinating coconut because the female flowers at the maturity stage are open to the environment and consequently to alien pollen that are transmitted via insects or winds. In general, stronger within population genetic structure has been observed in species that use insects and animals for pollination and dispersal (Hamrick et al. 1992) due to the out breeding that occurs. Yet, the deep genetic structure observed in SLYD coconut population cannot be attributed to cross pollination in a single generation alone. The smaller percentage of cross pollination that can happen in any dwarf population in a single generation may result in a certain proportion of heterozygosity. Such genetic heterozygosity needs to be carried down a few generations for them to get fixed and result in a genetic structure within population.

It also can be hypothesised that during the evolution process contribution of alleles from other coconut varieties such as local talls, intermediate varieties such as King coconut in Sri Lanka, by way of pollen donors, may have resulted in the population structure observed in SLYD. Such alien alleles that were mixed by cross pollination may have been fixed due to the naturally self-pollinating breeding behaviour of SLYD giving rise to the tall-like and mixed character groups of coconuts within the SLYD. The portion of outcrossing that may have occurred between the pure dwarf and tall-like group of coconuts bearing yellow-coloured nuts may have resulted in the category of palms included in the mixed genetic group identified in the current study. The palms in the mixed group can be expected to fix their characteristics and be included in either of the DL or TL group or even form new morphological groups upon continuous self-pollination. Accordingly, this population provides an example for the formation of new morphological groups during evolution even in a naturally self-pollinating tree crop as has been shown in this study.

In coconut, most dwarf varieties are extremely uniform as a result of not only their flowering pattern but also due to a strict stabilizing selection at the nursery stage. The policy adopted here in conserving yellow dwarf diversity was different aiming at conserving as much diversity as possible. The original population used for conservation may already have had a fairly high diversity level for a dwarf and it has expressed in the current progeny by further fixation of characters. While the method of conservation using open pollinated seednuts may be detrimental to conserve the existing diversity intact, it provides a practical method of creating morphological diversity by the formation of new allelic combinations. Also, since this population consists of dwarf, mixed (intermediate) and tall-like morphologies, it could serve as a useful model in studying the dwarfism in coconut.

The occurrence of population structure within dwarf coconut populations can be common to other colour forms (green, red and brown) of dwarf coconuts in the world as well. However, it can be hypothesised that the resulting

morphological variation may be masked or have become less conspicuous in green dwarfs due to the green colour of the nuts of resulting morphological groups. The majority of tall coconuts in the world bear green colour nuts and any morphological group arising by cross pollination of green and dwarf and tall coconuts would bear green colour nuts complicating the differentiation of these new morphological groups from the tall coconuts bearing green colour nuts.

The numbers of populations of brown and red dwarfs are very small and therefore this variation may not have been observed due to the limited scale of their occurrence.

A previous study reported the formation of new morphological groups in Malayan yellow dwarf (MYD) for resistance to lethal yellowing (Lebrun et al. 2008). The population differentiation of MYD for lethal yellowing may be due to evolution of both the population itself and/or the causal organism of the disease. The current study reports the formation of new morphological groups during the evolution of the plant population alone. There should be sufficiently large dwarf populations grown in close proximity for the development of a population structure. However, the dwarf coconuts are less famous in the world in plantation scale and therefore opportunities for the development of a within population genetic structure are minimal which explains the lack of documental evidence in the scientific literature.

Conclusions

The genetic structure revealed has been unexpected in a dwarf population as per the known characteristics of dwarf coconuts in the world. The information generated in the current study has been used to distinguish pure Sri Lankan yellow dwarf coconut. The possibility for the development of such a genetic structure and new genotypes during the evolutionary process has also been explained.

The information revealed in the current study will facilitate more effective utilization of dwarf germplasm in breeding programmes. This study has further demonstrated the fixation of new morphologies during the conservation if open pollinated progeny is used to derive the subsequent generation. Therefore deriving of planting material by artificial self-pollination is recommended for preservation of genetic uniqueness in dwarf coconuts in rejuvenation of field gene banks and for developing purelines for breeding purposes.

Acknowledgments This research was funded by the National Research Council of Sri Lanka under grant no. 11-042. Authors wish to thank Mrs. W. B. S. Fernando of the Coconut Research Institute of Sri Lanka for assistance provided in extraction of DNA.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

Data archiving statement Genotypic data will be submitted to database DRYAD if the manuscript is accepted for publication in the tree genetics and genomes prior to publication.

References

Anolles GC, Petter MG (1994) Staining nucleic acids with silver: an alternative to radio isotopic and fluorescent labelling. Promega Notes Magazine 45:13

Baudouin L, Lebrun P (2002) The development of a microsatellite kit and dedicated software for use with coconuts. Burotrop bull 17:16–20

Dasanayaka PN, Everard JMDT, Karunanayaka EH, Nanadasa HG (2009) Analysis of coconut (*Cocos nucifera* L.) diversity using microsatellite markers with emphasis on management and utilization of genetic resources. Journal of National Science Foundation of Sri Lanka 37(2):99–109

Doyle JJ, Doyle JL (1990) Isolation of plant DNA from fresh tissues. Focus 12:13–15

Gunn BF, Baudouin L, Olsen KM (2011) Independent origins of cultivated coconut (*Cocos nucifera* L.) in the old world tropics. PLoS One 6(6):e21143

Hamrick JL, Godt MJW, Sherman-Broyles SL (1992) Factors influencing levels of genetic diversity in woody plant species. New For 6: 95–124

Lebrun P, Baudouin L, Myrie W, Burger A, Dollet M (2008) Recent lethal yellowing outbreak: why is the Malayan yellow dwarf coconut no longer resistant in Jamaica? Tree Genet Genomes 4(1):125–131

Liu K, Muse SV (2005) PowerMarker; integrated analysis environment for genetic marker data. Bioinformatics 21(9):2128–2129

Meerow AW, Wisser RJ, Brown JS, 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 and Appl Genet 106:715–726

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

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. and Appl. Genet. 96:545–550

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 43:15–21

Perera L, Russell JR, Provan J, Powell W (2003) Studying genetic relationships among coconut varieties/populations using microsatellite markers. Euphytica 132:121–128

Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. Genetics 155:945–959

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

Rosenberg NA (2004) DISTRUCT: a program for the graphical display of population structure. Mol Ecol Notes 4:137–138

Weising K, Kahl G (1997) Hybridization-based microsatellite fingerprinting of plants and fungi. In: Casteno-Anolles G, Gresshoff PM (eds) DNA markers; protocols, applications and overviews. Wiley-VCH, p 27–53