Cord 2016, **32** (2)

Genetic Diversity Evaluation of MARDI's Coconut (*Cocos nucifera* L.) Germplasm using Simple Sequence Repeats

H.N. Khairun[1*], J. Azwan[1], A.R. Shahril[1], G.K. Sentoor[3], N. Siti[1], M.Y. Fairuz[1], H. Habibuddin[2] and A. Norzihan[1]

Abstract

A total of 18 simple sequence repeat (SSR) markers have been optimized and used to genotype coconut (*Cocos nucifera* L.). These markers were used to genotype 23 coconut varieties that were maintained in MARDI's germplasm collection. Fifteen SSR loci were polymorphic markers while the remaining three SSR were monomorphic. The number of alleles ranged from 2 to 19 with a mean number of 8.53 per locus. The expected heterozygosity values in each variety ranged from 0.07 to 0.61, with an average value of 0.52. Several varieties could be successfully differentiated by using these 15 SSR markers such as Cameroon Red Dwarf x West African Tall hybrid, Mawa hybrid, Malayan Tall x Cameroon Red Dwarf hybrid, Malayan Red Dwarf x Rotuman hybrid, Catigan, Pandan and Laguna. A UPGMA Dendogram showed Niu Damu as an outlying group with high dissimilarity from all other varieties. Analyses using the STRUCTURE software showed all 23 varieties to be clustered into 21 genotypic groups. This new information will greatly contribute towards characterization of the MARDI's coconut germplasm collection and to develop a SSR tool for the identification of new coconut varieties in Malaysia.

Keywords: *Cocos nucifera,* Simple sequence repeat (SSR), Genetic diversity, MARDI coconut germplasm

Introduction

Coconut (*Cocos nucifera* L) is placed in the Arecaceae family and the sub-family Cocoideae which has 27 genera (Teulat et al. 2000). In Malaysia coconut is distributed mainly in coastal regions between 20° N and 20° S, from sea level to 1,000 m altitude. Coconut has 32 chromosomes (16 pairs) and can be divided into two morphological types, namely Tall and Dwarf. The Tall is generally cross-fertilized while the Dwarf is self-fertilized. In the natural state, the Dwarf exists very rarely. The morphological diversity and geographical distribution of the coconut has led to the identification of more than 600 putative diverse coconut varieties (Coconut Genetic Resources Database Ver.4/COGENT/IPGRI).

[1]Centre for Molecular Marker Discovery and Validation (CMDV), Technical Service and Laboratory Centre, MARDI 43400 Serdang, Selangor, Malaysia.

[2]Rice Research Centre, MARDI Headquarters, 43400 Serdang, Selangor, Malaysia.

[3]Soil Science Centre, MARDI Headquarters, 43400 Serdang, Selangor, Malaysia.

*Corresponding author

E-mail: hairin@mardi.gov.my

In Malaysia, coconut ranks as the fourth most important crop in terms of area planted, after *Elaeis guineensis*, *Hevea brasiliensis* and *Oryza sativa* (Yeow et al, 2009). The coconut planted area has drastically declined as reported by FAO (2007) with the recorded planting area of 409,384 ha in 1981, 248,380 ha in 1995 and 172,000 ha in 2007. The declining coconut planted area is due to competition for land, oil palm planting and infrastructure development in the country. The Malaysian government, primarily through MARDI, collaborated with a number of international coconut R&D organizations, particularly the International Plant Genetic Resources Institute/International Coconut Genetic Resources Network (IPGRI/ COGENT) to undertake research and development activities. MARDI has identified74 coconut varieties (FAO 2007) which were evaluated and characterized based on their morphology. As morphology characterization is largely influenced by environmental factors, this collections would probably have duplicates or types that are genetically very similar. Therefore, a reliable assessment of the genetic relationships among varieties and accurate estimation of the genetic diversity present in the coconut is a pre-requisite for sustaining future coconut breeding and genetic resources conservation programmes. Thus, assessment techniques that are independent of environmental influences are important to plant breeders.

Evaluation of coconut phenotypic traits is direct, cost effective and easy to do but they are influenced by environmental factors (Perera et al., 2003). In this regard, molecular marker techniques are advantageous over the morphological measurements for the assessment of genetic diversity as they directly reflect variations in the DNA sequence which are free from environmental influences and provide additional information. Among many molecular marker techniques currently available, microsatellites or SSRs are very useful for assessing genetic diversity and genetic relationships in plants as they are highly polymorphic, co-dominant, very informative, PCR based, highly repeatable and can be integrated with high throughput instrumentation. In this study, we have demonstrated the application of SSR markers in assessing the genetic diversity present in 23 varieties of MARDI's coconut germplasm collection. The information generated will provide a deeper insight on the collections and thus will help MARDI on managing and protecting the coconut varieties and to develop a SSR panel for characterizing and identification of potentially new local coconut varieties.

Materials and Methods

Plant material

Young leaves from a total of 383 individual plants, belonging to 23 coconut varieties grown in MARDI Hilir Perak were collected (Table 1) and stored in silica gel until further use for genomic DNA extraction.

Genomic DNA extraction and DNA quality control

Genomic DNA was extracted using modified CTAB DNA extraction method. About 1 g of young coconut leaves were placed into a 96-well plate together with small stainless balls and ground to a powder using a TissueLyser (Qiagen). To this was added 600 uL of an extraction buffer (final concentration; PVP 2%, DIECA 4 mM, ascorbic acid 5 mM, NaCl 1.4 M, Tris-HCL(pH 8.0) 100 mM and EDTA 20 mM). The mixture was then incubated at 65°C for 1 hour with intermittent mixing. To this, the same volume of isopropanol was added and the mixture homogenized by inverting the plates. The mixture was centrifuged at 5,500 rpm for 15 minutes. The supernatant was then transferred into an equal volume of isopropanol and centrifuged at 5,500 rpm for a further 15 minutes. The DNA was air-dried and suspended in 50 µLTris EDTA (TE) buffer. The DNA concentration was quantified using a Thermo Labsystems Fluoroskan Ascent ™ (Thermo Scientific, USA) and DNA integrity determined on a 0.8% agarose gel.

SSR Genotyping

Thirty SSR markers from published studies (Baudouin et al., 2006, Rivera et al., 1999, Meerow et al., 2003 and Dasanayaka et al.,

Table 1. The 23 coconut varieties from the MARDI's germplasm collection used in this study

No	Type	Variety	Sample number
1	Tall	Tagnanan	9
2	Tall	Catigan	20
3	Tall	West African Tall (WAT)	15
4	Tall	Fiji	13
5	Tall	Laguna	20
6	Tall	Niu Leka	20
7	Tall	Niu Damu	20
8	Tall	Niu Kitu	7
9	Tall	Pinggan-Pinggan	18
10	Dwarf	Malayan Yellow Dwarf (MYD)	20
11	Dwarf	Malayan Red Dwarf (MRD)	12
12	Dwarf	Malayan Brown Dwarf (MBD)	12
13	Dwarf	Malayan Green Dwarf (MGD)	16
14	Hybrid	Mawa	16
15	Dwarf	Pandan	16
16	Hybrid	Matag	10
17	Hybrid	Maren	19
18	Hybrid	MRD X Kar Kar	20
19	Hybrid	MRD X Rotuman Tall (RT)	20
20	Hybrid	MRD X Markham Valley Tall (MT)	20
21	Hybrid	Cameroon Red Dwarf (CRD) X Rennel Tall	20
22	Hybrid	Malayan Tall (MT) X CRD	20
23	Hybrid	CRD X WAT	20
	Total		383

2009) were used to asses 23 coconut varieties from MARDI's germplasms collection. Each SSR markers were tagged with M13 sequence as an adaptor to ligate with the fluorescent dye. PCR amplification was carried out in a total reaction volume of 10 µL consisting of 1.0 µL of template DNA (approximately 40 ng μL^{-1}), 1 uL PCR reaction buffer (1.5-3.0 mM $MgCl_2$), 2 mM dNTPs mixtures, 10 µmol of each primer set, 5 µmol of a fluorescent dye (FAM/VIC/NED and PET) and 1 Unit *Taq* DNA polymerase (Invitrogen, USA). Amplification was performed using a Peltier Thermal Cycler, DNA Engine Tetrad 2 (Biorad, USA). The PCR profile was an initial denaturation at 94°C for 5 minutes, followed by 34 cycles of a second denaturation at 94°C for 30 s, annealing (see Table 2**)** for 45 s, and 72°C for 45 s extension, and finally one cycle of 72°C for 7 minutes. The PCR products were genotyped using the 3730xl DNA Analyzer (Applied Biosystem, USA).

Scoring and Data Analyses

The fragment analysis files were analysed using the GeneMapper 5.0 software (Applied Biosystems). GS500LIZ was used as the standard control. Allele peaks in the electropherograms were scored and analyzed as by the method of Arif et al. (2010). The peaks scored were imported into a Microsoft Excel file and the data pre-analysed using MicroChecker v2.2.3 (Van Oosterhout et al., 2004) to check for the presence of null alleles, large allele dropouts and scoring errors due to stutter peaks. Passed data, after MicroChecker analysis, was converted into input data file using CONVERT 1.2 (Glaubitz, 2004) for further analyses. POPGENE32 version 1.31 was used to calculate the number of alleles, allele frequencies of each varieties and the genetic distance between varieties. The NTSYSpc software was used to construct dendograms using the UPGMA method based on the genetic distance produced from the POPGENE32 software. The program STRUCTURE was used to investigate the varietal structure (Pritchard et al.2000; Falush et al. 2003, Rohlf, F. J. 2008).

Results and Discussion

A total of 383 individual samples representing 23 varieties in the MARDI's coconut germplasm collection were successfully analyzed to better understand their genetic diversity and the genetic relationships among those varieties. Thirty SSR markers were obtained from the public domain, however only 25 SSR markers were successfully optimized. Out of these markers, only 15 SSR markers were able to detect polymorphisms and three loci to be monomorphic. These SSR markers generated amplicon sizes ranging from 94 to 310 bp. The number of allele ranged from 2 to 19 with a mean number of alleles per locus at 8.53. Locus CN11E10 showed the highest effective number of alleles even though the number of alleles detected for CN11E10 (12 alleles) was lower than CN11E6 (19 alleles) (Table 2). Based on this observation, locus CN11E10 showed a high discriminatory power as compared to the other loci (Table 3). Garza and Williamson (G/W) Index (Garza and Williamson, 2001) inferred, varieties that have values lower than the threshold bottleneck of 0.68 showed evidence for recent reduction in population sizes. Thus, with this assumption all varieties except for the varieties Matag, Catigan, MYD, MRD, MBD, MGD and Pandan experienced reduction in population sizes (Table 4). While the WAT varieties showed the most severe reduction in population size base on their G/W value. This G/W index suggested that bottlenecks had occurred in most of the varieties in the MARDI germplasm collection. One action needed to be taken, to ensure the genetic diversity was maintained in the germplasm collection, was to protect the varieties from disease wipe out and environmental stress. The results also showed there were no useful and polymorphic loci in the Matag population. This could be explained by the phenotypic variation in the Matag population itself since there were three husk colours to the Matag fruit found in Malaysia (green, brown and yellow). This situation could also be explained by the low number of usable loci and polymorphic loci in the Tagnanan population which is one of the Matag's parent.

Table 2. 15 microsatellite markers used in this study

Primer	*Primer sequences (5'-3')*	*Repeat Motive*	*Repeat Type*	*No. of alleles*	*Product size*	*T^a (°c)*	*H^o*	*H^e*
CACO8	F: ATCACCCCAATACAAGGACA R: AATTCTATGGTCCACCCACA	$(AG)^{10}(CA)^9$	Di	7	190	56.7	0.2305	0.459
CAC20	F: CTCATGAACCAAACGTTATA R: CATCATATACATACATGCAACA	$(CA)^{19}$	Di	9	124-133	56.7	0.4336	0.5463
CAC21	F: AATTGTGTGACACGTAGCC R: GCATAACTCTTTCATAAGGGA	$(CA)^{11}$	Di	4	149-151	50.5	0.3832	0.5196
CAC38	F: ACCCTACTTCTAACTGTTCACTC R: CAGCTTGATAAATATCATCCAT	$(CA)^{13}(CT)^{17}$	Di	11	155	50.5	0.4329	0.5952
CN11E6	F: TACTTAGGCAACGTTCCATTC R: TAACCAGAAAGCAAAAAGATT	$(CT)^{21}$	Di	19	85-128	50.5	0.4536	0.5114
CN11E10	F: AGAGAGAGTAAATGGGTAAGT R: CCCTTTCATTTTTCCTTATTC	$(GT)^{22}(GA)^{14}$	Di	12	99-151	50.5	0.3489	0.8218
CnCir 126	F: TAATGACCTCTGCCG R: CCTGATTGGGTGTCTAT	$(AAAG)^5$ $(GA)^{10}$	Tetra/ Di	3	179-190	53.4	0.1432	0.1775
CnCir 147	F: TTTCTCACCAACAAATAAAC R: CTTGTGTGTTAGGGTCATC	$(CT)^9(GT)^2(CT)^6$	Di	12	205-233	48.2	0.5086	0.687
CnCir 215	F: TACCACCTAAAGAGGAATG R: AGTATCTGGGTTTGGCT	$(CT)^{17}$	Di	7	114-176	53.4	0.5978	0.6371
CnCir F3	F: CCCTACTACTCCCTCAT R: TGCCTAGTCAATCATAC	$(TC)^{17}(CA)^{16}$	Di	8	136-199	48.2	0.338	0.579
CnCirH4	F: TTAGATCTCCTCCCAAAG R: ATCGAAAGAACAGTCACG	$(CT)^8(CA)^5(CGCA)^6$	Tetra/ Di	4	218–236	64.6	0.3066	0.3632
CnCir I4	F: TCCTAGTGCTTATGCTTGAC R: TTGATGGTTTGATGTTGAA	$(CT)^9(CA)^8$	Di	2	285-292	50.5	0.0667	0.0645
CNZ01	F: ATGATGATCTCTGGTTAGGCT R: AAATGAGGGTTTGGAAGGATT	$(CT)^{15}(CA)^9$	Di	8	109-131	56.7	0.3083	0.5608
CNZ43	F: TCTTCATTTGATGAGAATGCT R: ACCGTATTCACCATTCTAACA	$(GA)^{21}$	Di	12	197	59.6	0.4776	0.6254
CNZ44	F: CATCAGTTCCACTCTCATTTC R: CAACAAAAGACATAGGTGGTC	$(GA)^{15}$	Di	10	165	59.6	0.5169	0.6364

T^a = annealing temperature; H^e = expected heterozygosity; H^o = observed heterozygosity; PIC = Polymorphic Information Content

Table 3. Number of alleles for each locus

No.	Locus	Sample size	na*	ne*
1	C5	642	7	1.846
2	C7	738	9	2.2005
3	C8	736	4	2.0785
4	C10	656	11	2.4648
5	C27	754	19	2.0437
6	C28	642	12	5.5704
7	C30	754	3	1.2155
8	C31	696	12	3.1844
9	C32	736	3	1.2155
10	C34	716	7	2.7487
11	C35	698	12	3.1844
12	C36	750	4	1.5691
13	C37	720	8	2.3705
14	C45	624	12	2.6624
15	C46	592	10	2.7425
		Mean	8.5333	2.4026
		St. dev	4.4218	1.0501
	*na	Observed number of allele ,		
	*ne	Effective number of allele		

An Unweighted Pair Group Method with Arithmetic Mean (UPGMA) dendogram based on the Jaccard dissimilarity matrix was constructed using the NTSYS software package to investigate the relationships of these coconut varieties. The complete linkage method was carried out on the taxonomic distance matrix to establish the relationship among the varieties. The dendrogram showed that the varieties were grouped into 4 major clusters and Niu Damu was out group with high dissimilarity from the other varieties (Figs 1). However the dendogram for individual clustering showed certain markers were able to differentiate some of the varieties such as Catigan, CRD x WAT, Mawa, Pandan, Laguna, MRD x RT and MT x CRD. Those markers will further validate using more samples for development fingerprinting panel.

A schematic representation of the genetic structure of all the individuals under this studied was represented in figure 2. The STRUCTURE software uses allele frequencies from what was assumed to be potentially unlinked loci and uses these figures to identify which set of genes belong to which variety. Based on the alleles frequencies the varieties was set (K), but K does not represent the number of varieties sample and it can be anywhere from 1 to the number of varieties collected. Thus, this software will theoretically indicate the total number of varieties that represent the data being analyzed. Thereafter the individuals were assigned to their

Table 4. Level of heterozygosity of the 23 coconut varieties

No	Varieties	Observed Heterozygosity	Expected Heterozygosity	G/W index	Usable Loci	Polymorphic Loci
1	Matag	0.27881	0.2995	0.7113	0	0
2	Tagnanan	0.30963	0.44035	0.6782	2	1
3	MRD X Kar kar	0.32299	0.33893	0.5495	10	10
4	MRD X RT	0.39253	0.41106	0.5932	10	8
5	MRD X Markham Valley Tall (MT)	0.37451	0.33171	0.5992	12	12
6	Catigan	0.08827	0.20291	0.7346	12	7
7	WAT	0.26256	0.30194	0.4563	9	8
8	CRD X Rennel Tall	0.6181	0.4707	0.5678	9	9
9	MT X MRD	0.49101	0.43753	0.6419	10	9
10	CRD X WAT	0.739	0.48209	0.5261	10	9
11	Fiji	0.47	0.61638	0.6148	9	8
12	Maren	0.50123	0.523	0.5983	7	7
13	Laguna	0.57047	0.48754	0.5573	12	12
14	Niu Leka	0.42078	0.56142	0.5894	11	11
15	Niu Damu	0.3567	0.50196	0.6395	12	11
16	Niu Kitu	0.38794	0.45484	0.6675	10	7
17	Pinggan-Pinggan	0.38303	0.5419	0.6784	5	4
18	MYD	0.03303	0.07624	0.9087	12	4
19	MRD	0.02222	0.13024	0.8806	8	2
20	MBD	0.05667	0.14493	0.9028	9	2
21	MGD	0.09702	0.15837	0.8289	10	2
22	Mawa	0.66999	0.45729	0.5062	8	7
23	Pandan	0.12635	0.10959	0.8889	8	5

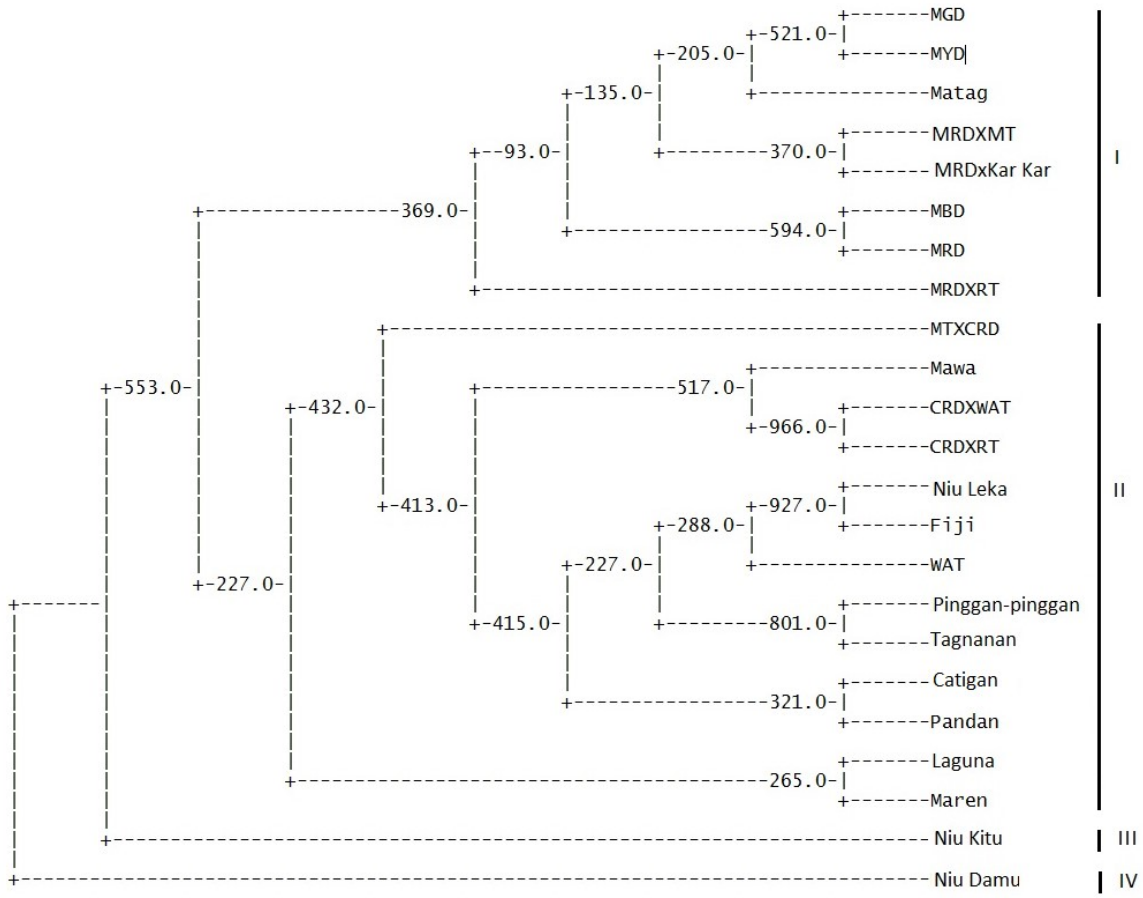

Figure 1. UPGMA dendogram for 23 coconut varieties were group into 4 clusters

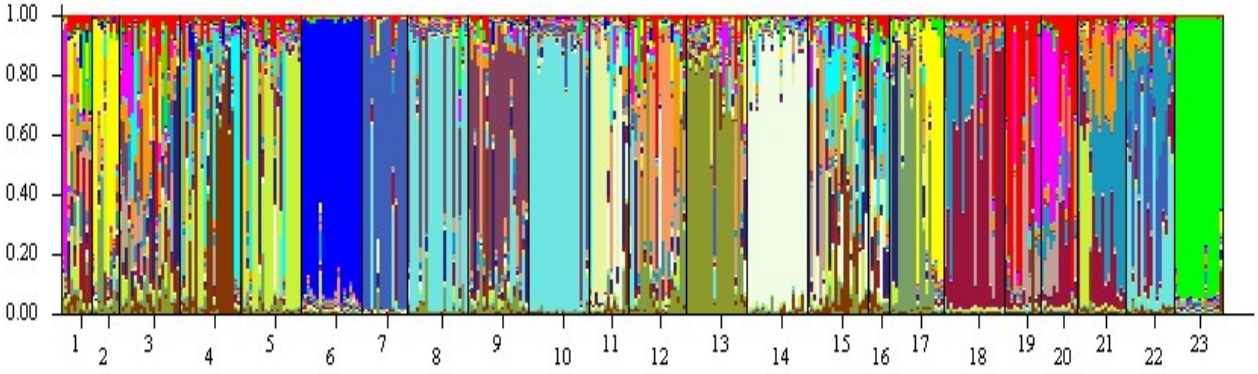

Figure 2. Results of the Cluster Analysis generated by the *STRUCTURE* software. Different colours indicate different genetic clusters. Each column represents an individual whereby the height of the column segment shows the probability of assignment of the individual *Cocos nucifera* to a genetic cluster. 1=Matag, 2=Tagnanan, 3=MRD X Kar kar, 4= MRD X RT, 5= MRD X MT, 6= Catigan, 7= WAT, 8= CRD X RT, 9= MT X CRD, 10= CRD X WAT, 11= Fiji, 12= Maren, 13= Laguna, 14= Niu Leka, 15= Niu Damu, 16= Niu kitu, 17= Pinggan-pinggan , 18= MYD, 19=MRD, 20= MBD, 21= MGD, 22= Mawa and 23= Pandan

significant K varieties (Pritchard et al, 2000). The log probabilities [Ln P(D)] associated with different numbers of genetic clusters (K), calculated from Bayesian clustering analysis based on 383 individuals with 15 loci showed the highest value of K = 21[Ln P(D) = -15198.3]. According to the degree of admixture (alpha value), K =21 gives the nearest value to zero which indicates that the most appropriate number of genetic groups assigned for the given data set was 21. The result showed that Tagnanan and Matag shared similar genotypic frequency. This also could be seen for varieties CRD x RT and CRD x WAT. The minimal number of coloured lines in a few varieties showed a high possibility inbreeding occurring in those varieties. A Genetic differentiation (F_{st}) values with the $p < 0.05$ was the cut-off point of significant varietal differentiation. The data demonstrated significant differentiation that ranged from 0.0020 to 0.3381. The highest divergence value was between Catigan, CRD x WAT, MAWA and Pandan varieties while the lowest was between the Matag and Tagnanan varieties.

Conclusion

Genetic diversity studies on 23 coconut varieties in the MARDI's germplasm collection identified 15 SSR markers that showed polymorphism among the individual samples. Locus CN11E10 showed the highest effective number of alleles indicating a higher differentiating power over the other loci. Analysis using the STRUCTURE software classified the 23 coconut varieties into 21 genotypic clusters. These results provide useful information on the genetic diversity of the collection and will assist breeders in planning for future selections programmes using these coconut varieties.

References

Arif, I. A. et al. (2010) Interpretation of electrophoretograms of seven microsatellite loci to determine the genetic diversity of the Arabian Oryx. Genetics and Molecular Research, 9(1), 259-265.

Baudouin, L. et al. (2006) QTL analysis of fruit components in the progeny of a Rennell Island Tall coconut (Cocos nucifera L.) individual. Theoretical and applied genetics, 112(2), 258-268.

Dasanayaka, P. N. et al. (2009) Analysis of coconut (Cocos nucifera L.) diversity using microsatellite markers with emphasis on management and utilisation of genetic resources. Journal of the National Science Foundation of Sri Lanka, 37(2).

Falush, D. et al. (2003) Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies. *Genetics*, 164(4), 1567-1587.

Garza, J. C., & Williamson, E. G. (2001) Detection of reduction in population size using data from microsatellite loci. *Molecular ecology*, 10(2), 305-318.

Glaubitz, J. C. (2004) Convert: A user-friendly program to reformat diploid genotypic data for commonly used population genetic software packages. *Molecular Ecology Notes*, 4(2), 309-310.

Meerow, A. W. et al. (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*(4), 715-726.

Pritchard, J. K. et al. (2000) Inference of population structure using multilocus genotype data. Genetics, 155(2), 945-959.

Perera, L. et al. (2003) Studying genetic relationships among coconut varieties/populations using microsatellite markers. Euphytica,132(1), 121-128.

Rivera, R. et al. (1999) Isolation and characterization of polymorphic microsatellites in *Cocos nucifera* L. *Genome*, *42*(4), 668-675.