Theor Appl Genet (2000) 100:764–771

B. Teulat · C. Aldam · R. Trehin · P. Lebrun
J.H.A. Barker · G.M. Arnold · A. Karp · L. Baudouin
F. Rognon

An analysis of genetic diversity in coconut (*Cocos nucifera*) populations from across the geographic range using sequence-tagged microsatellites (SSRs) and AFLPs

Received: 27 October 1998 / Accepted: 31 March 1999

Abstract Genetic diversity in 31 individuals from 14 coconut populations across the entire geographic range (2–3 individuals per population) was assessed using sequence-tagged microsatellites (or simple sequence repeats, SSRs) and amplified fragment length polymorphism (AFLP). From the 39 SSR primer sets tested, only two gave patterns that could not be scored and used in the data analysis. The remainder included five SSRs that gave double-locus profiles in which one locus could still be scored separately. The 37 SSRs revealed between 2 and 16 alleles per locus and a total of 339 alleles in the 14 populations. Gene diversity ($D = 1-\Sigma p_i^2$) ranged from 0.47 to 0.90. Two of the four Dwarf populations were homozygous at all 37 loci, which is consistent with their autogamous (self-fertilising) reproduction. One Dwarf population was heterozygous at one locus but the other (Niu Leka Dwarf), which is known to be cross-pollinating, showed high levels of heterozygosity. Generally, diversity was higher in populations from the South Pacific and South East Asia. Three SSR loci (CNZ46, CN2A5, CN11E6) gave distinct genotypes for all but two populations. The East African populations had higher heterozygosities than those from West Africa, and the populations from Tonga and Fiji generally had distinct alleles from those of the South Pacific. AFLP analysis with 12 primer combinations gave a total of 1106 bands, of which 303 were polymorphic (27%). Similarity matrices were constructed from the two data sets using the proportion of shared alleles for SSRs and a Jaccard coefficient for AFLPs. In each case cluster and principal coordinates analyses were performed, with the resultant dendrograms and plots revealing similar relationships among the populations for both approaches. There was generally a good separation of populations, and phenetic relationships were in agreement with those previously shown by RFLPs. The use of SSRs and AFLPs in genetic-diversity analysis for the establishment of germplasm collections is discussed.

Communicated by J.W. Snape

B. Teulat · F. Rognon
Burotrop, Agropolis International, Avenue Agropolis,
34394, Montpellier, Cedex 5, France

C. Aldam · R. Trehin · J.H.A. Barker · G.M. Arnold · A. Karp (✉)
IACR-Long Ashton Research Station,
Department of Agricultural Sciences,
University of Bristol, Long Ashton, Bristol, BS 41 9AF, UK
e-mail: angela.karp@bbsrc.ac.uk,
Tel.: +44 1275 392181

P. Lebrun · L. Baudouin
CIRAD, Amélioration Cocotier, BP 5035,
34032 Montpellier Cedex 01, France

Introduction

Coconut (*Cocos nucifera* L.) is one of the key plantation crops of the tropics, where its food and industrial products, such as coconut oil, copra and desiccated coconut, play an important role in the economies of many developing countries. Almost every part of the coconut tree is used either to generate income or to meet the food requirements of rural communities. In subsistence and semi-subsistence farming systems, coconut provides reliable insurance as a food supply even when other crops fail. Coconut can be cultivated in sandy coastal regions since it grows well in areas of low rainfall where soil drainage is good, or where there is a good supply of ground water, and can tolerate higher soil salinity than many other crops. Its current distribution spans the coastal areas between 20° north and 20° south of the equator and ranges in altitude from sea level to 1000 m. Before its domestication in the Indo-Pacific region, it is thought that the ancestors of coconut may have come from the Far East and Pacific regions (Harries 1990), but there are conflicting theories regarding the origin and domestication of coconut (Whitehead 1976; Harries 1978).

C. nucifera ($2n = 2x = 32$) is a member of the monocotyledonous family Aracaceae (Palmaceae). It is the only species of the genus *Cocos* belonging to the subfamily Cocoideae which includes 27 genera and 600 species. The geographic spread of coconut and the diversity of the palm in growth habit, fruit shape and colour, have led

to the identification of a large number of varieties. There are two main types: Tall coconuts are fast growing and predominantly allogamous (cross-fertilising) whilst, with the exception of Niu Leka, Dwarfs have a reduced growth habit and are mostly autogamous (self-fertilising). The latter varieties are thus considered to be fixed lines, while the former constitute polymorphic populations and cultivars are either populations maintained under natural pollination or hybrids between populations. Some 900 accessions are currently recognised by the International Coconut Genetic Resources Network (COGENT). Coconut cultivation is currently in relative decline in many countries, however, due to strong competition from other oil crops such as oil palm, low productivity, increased demand for timber and devastating diseases. Germplasm collections that contain significant amounts of genetic diversity within and among coconut populations are essential for an effective crop improvement programme. COGENT is planning to collect threatened and useful germplasm to augment the collections, with a view to establishing regional gene banks. In this endeavour, sensitive and reliable techniques for assessing genetic diversity will be an important adjunct for the establishment of effective collecting and gene-banking strategies.

Characterisation and evaluation of coconut populations based on morphological and agronomic traits is time-consuming and labour-intensive and provides only a simplified picture of diversity (Akpan 1994; Sugimura et al. 1997). Variations in isozymes (Carpio 1982), fruit components (Ashburner et al. 1997 a) and carotenoid differences (Fernando et al. 1997) can be limited in discriminatory capacity. More recently, coconut germplasm has been characterised using primers specific for copia-like *Eco*RI elements (Rhode et al. 1995), randomly amplified polymorphic DNA (RAPD) (Ashburner et al. 1997 b), restriction fragment length polymorphism (RFLP) (Leburn et al. 1998 a, b) and amplified fragment length polymorphism (AFLP) (Perera et al. 1998). Such DNA-based assays are attractive because of the immense number of ‘characters’ they reveal and their capacity to be performed with greater speed and efficiency, at all stages during the life cycle and independently of the environment.

Of the DNA-based methods that could be used to assess coconut germplasm, only a few are suitable for application in networked efforts spanning several continents, consistent with the COGENT strategy. Results must be reproducible by different laboratories and analysed using standardised scoring and analytical methods. Data must be easily entered into databases. Assays should be applicable for high throughput and necessary materials should be easily exchangeable. Since ‘‘populations” need to be analysed, co-dominant markers are preferable and high polymorphism rates may be required to distinguish closely related genotypes.

Microsatellites (or simple sequence repeats, SSRs), have many features which help meet these criteria (see Gupta et al. 1996; Powell et al. 1996 a). In consideration of their potential use in coconut, we previously isolated a set of 39 highly polymorphic microsatellites from the cultivar Tagnanan Tall which were tested on a limited range of 20 coconut cultivars from the Philippines (Rivera et al. 1999). Here, we report on the use of these SSRs to evaluate genetic diversity in 14 populations from across the geographic range. These populations have previously been characterised using RFLPs (Lebrun et al. 1998 a, b), thus permitting a comparison of the two marker types. Furthermore, since AFLPs have recently been recommended for high-resolution analysis of genetic diversity in coconut (Perera et al. 1998), we analysed the same 14 populations using AFLPs with 12 different primer combinations.

Materials and methods

Materials and DNA extraction

Ten Tall and four Dwarf populations from across the geographic range were selected and represented by 31 individual palms (at least two per population) (Table 1). All populations are maintained at Marc Delorme Research Station (CNRA, Côte d’Ivoire). DNA was extracted at CIRAD (Montpellier, France) from lyophilised leaves of frond number one using a CTAB method as described in Lebrun et al. (1998 a).

PCR assay for SSR primer sets and SSR genotyping

PCR amplification was performed at IACR-Long Ashton Research Station (Bristol, UK) in 0.5 ml thin-walled tubes (NBL Gene Sciences Ltd.) using a Perkin Elmer 9600. A list of the primer sequences used for the reactions is given in Table 2. In each case, the forward primer was radiolabelled with [γ-^{33}P]ATP by T4 kinase. The PCR reaction mixture (12.5 µl) contained 5 ng template DNA, 25 ng forward primer, 25 ng reverse primer, 100 µM of each dNTP (Promega), 0.5 U *Taq* DNA polymerase (Gibco BRL), 20 mM Tris-HCl pH 8.4, 50 mM KCl and 1.5 mM $MgCl_2$. The PCR regime consisted of an initial denaturation (94°C for 2 min), 35 cycles each consisting of 40 s denaturation (94°C), 1 min annealing (54°C) and 2 min elongation (72°C). At the end of the final run, an extension period of 10 min at 72°C was included. For use in genotyping, the reactions were mixed with 5 µl of stop solution [98% formamide, containing 10 mM EDTA, 0.01% (w/v) xylene cyanol and 0.01% (w/v) Bromophenol Blue] and denatured at 94°C for 3 min. A 4 µl aliquot was loaded onto denaturing polyacrylamide gels (6% acrylamide/bisacrylamide 40:2, 7 M urea in 1 × TBE buffer) and electrophoresed at 55 W constant power for 2 h. The gels were sprayed with CTAB fixer (0.3% CTAB; 0.2 M KH_2PO_4) and dried using a gel drier (Bio-Rad) at 80°C for 30 min. The dried gels were exposed to Kodak Biomax MR-2 film for 18 h.

AFLPs

AFLP analysis was performed essentially as described in the original method by Zabeau and Vos (1993). The primer and adapter sequences were synthesised and desalted by Genosys Biotechnologies Inc.

*Pst*I adapter: 5′-biotin-CTCGTAGACTGCGTACATGCA-3′
3′- CATCTGACGCATGT-5′
*Mse*I adapter: 5′-GACGATGAGTCCTGAG-3′
3′- TACTCAGGACTCAT-5′

For selective PCR, 12 *Mse*I and *Pst*I primer combinations were used: *Mse*I+GAA and *Pst*I+AAA; *Mse*I+GAA and *Pst*I+AAC;

Table 1 List of the coconut populations analysed. The codes identify individuals from the same population. Cultivar names are also given

Code	Sample no. [a]	Region	Country	Cultivar
WAT4	1, 2	West Africa	Ivory Coast	West African Tall (Mensah)
MZT	8, 9	East Africa	Mozambique	Mozambique Tall
MLT	3, 4, 5	South East Asia	Malaysia	Malayan Tall
MYD	29, 30	South East Asia	Malaysia	Malayan Yellow Dwarf
BAYT	6, 7	South East Asia	Philippines	Baybay Tall
SLT	25, 26	South Asia	Sri Lanka	Sri Lankan Tall
ADOT	10, 11	South Asia	Indonesia	Andaman Tall
KIGD	12, 13	South Pacific	Kiribati	Kiribati Green Dwarf
MBD	23, 24	South Pacific	Papua New Guinea	Madang Brown Dwarf
KKT	21, 22	South Pacific	Papua New Guinea	Karkar Tall
TONT	14, 15	South Pacific	Tonga	Tonga Tall
RIT	16, 17, 18	South Pacific	Soloman Islands	Rennell Island Tall
NLAD	27, 28, 31	South Pacific	Fiji	Niu Leka Dwarf
PNT2	19, 20	Latin America	Panama	Panama Tall (Monagre)

[a] Note: the samples listed here are ordered in terms of geographic region. The numbering 1–31 refers to the arbitrary numbers given to the populations in the numerical series in which they were consistently loaded onto gels (see Figs. 1 and 3)

*Mse*I+GAA and *Pst*I+CC; *Mse*I+ACA and *Pst*I+AAA; *Mse*I+ACA and *Pst*I+AAC; *Mse*I+ACA and *Pst*I+AAG; *Mse*I+GAA and *Pst*I+CAG; *Mse*I+GAA and *Pst*I+CAT; *Mse*I+ACA and *Pst*I+CAA; *Mse*I+ACA and *Pst*I+CAC; *Mse*I+ACA and *Pst*I+CAG; *Mse*I+ACA and *Pst*I+CAT (where *Mse*I=5′-GATGAGTCCTGAG-TAA-3′ and *Pst*I=5′-ACTGCGTACATGCAG-3′). In each case, the *Mse*I primer was end-labelled using [γ–^{33}P]ATP (Amersham) and T4 polynucleotide kinase, according to Vos et al. (1995). For each primer combination, the PCR reactions were performed in a total volume of 10 µl containing 1 µl bead/DNA template, 10 mM Tris-HCl, 1.5 mM $MgCl_2$, 50 mM KCl, pH 8.5, 2.5 ng ^{33}P-labelled *Mse*I primer, 12.5 ng unlabelled *Mse*I primer, 15 ng *Pst*I primer and 0.5 U *Taq* DNA polymerase (Gibco BRL). Amplification was performed using a Perkin Elmer 9600 for 10 cycles with the following cycle profile: 40 s DNA denaturation step at 94°C, 60 s annealing step at 62°C and a 60 s extension step at 72°C. The PCR was then continued for an additional 25 cycles with a lower annealing temperature of 56°C. The final 72°C step was extended for 10 min. The amplification products were mixed with an equal volume of 2× AFLP loading buffer [98% formamide, 10 mM EDTA, 0.01% xylene cyanol and 0.01% Bromophenol Blue] and 4 µl of each sample was separated on a 4.5% denaturing acrylamide gel, electrophoresed at 55 W power for approximately 2.5 h. The gel was sprayed with CTAB fixer, dried and exposed to Kodak Biomax MR-2 film for 18 h.

Data analysis

Microsatellite loci were scored individually and the different alleles were recorded for each sample screened, with single bands taken to indicate the presence of two identical alleles. A similarity matrix was calculated from these data based on the proportion of shared alleles at each locus and meaned across all loci. Information on allele size (repeat number) was not used in the calculation of these similarities because of the complex nature of a number of the microsatellites. AFLP bands were scored manually as present (1) or absent (0), with only polymorphic bands being included in subsequent analyses. A similarity matrix was constructed using the Jaccard coefficient for the bands from different primer combinations together. Cluster analyses were performed on the similarity matrices using the unweighted pair group method with arithmetic averages (UPGMA) and dendrograms (phenograms) constructed from these analyses. Similarity matrices were also used as inputs into principal co-ordinates analyses (PCO), with the resultant scores for samples on the first two components plotted pairwise in each case. All analyses were conducted using the Genstat 5 statistical package (Genstat 5 Committee 1993). For SSRs, gene diversity ($D = 1-\Sigma p_i^2$) values were calculated according to Nei (1973).

Results and discussion

Thirty nine microsatellite primer sets were tested on 14 coconut populations from across the geographic range. An example of SSR polymorphism is shown in Fig. 1. The number of alleles, the allele size ranges and the gene diversity (D) obtained (Table 2) indicate that high levels of polymorphism were detected by the SSRs. An average of 9.14 alleles per locus was observed and D ranged from 0.47 to 0.90.

Two primer sets (CNZ13 and CNZ22) gave amplification profiles that could not be easily scored and were thus excluded from the analysis. In the case of CNZ13, extensive stutter bands were produced which meant that accurate scoring was difficult to achieve. Primer set CNZ22 gave a complex multilocus profile, which could not be interpreted. An additional five SSRs (CNZ09, CNZ42, CN2A5, CN2A4 and CN11E10) also gave multilocus profiles, but in these cases one locus was sufficiently separated to be scored easily. For a further four primers (CNZ01, CNZ24, CN1G4 and CN11E6), incomplete data sets were generated due to repeated problems with amplification in one or more samples. As a result of these problems, gene diversity was estimated for 33 SSRs only (Table 2).

The SSRs were informative in revealing genetic differences among the populations (Table 3). Three SSR loci (CNZ46, CN2A5, CN11E6) gave distinct genotypes for all but two populations tested, suggesting that very few SSRs would be needed to distinguish all possible populations of coconut. Some alleles were specific to one population in the samples studied: e.g. CNZ05 gave a distinct allele for TONT, CNZ05 and CNZ34 each gave a distinct allele for NLAD, and CNZ19 gave a distinct allele for ADOT. In several cases, all three samples of the same population contained different allelic combinations (e.g. CNZ02 for NLAD and CN1G4 for RIT) suggesting that SSRs may be a useful tool in identifying distinct individuals within populations. Within- and between-population variability appeared highest in the Tall

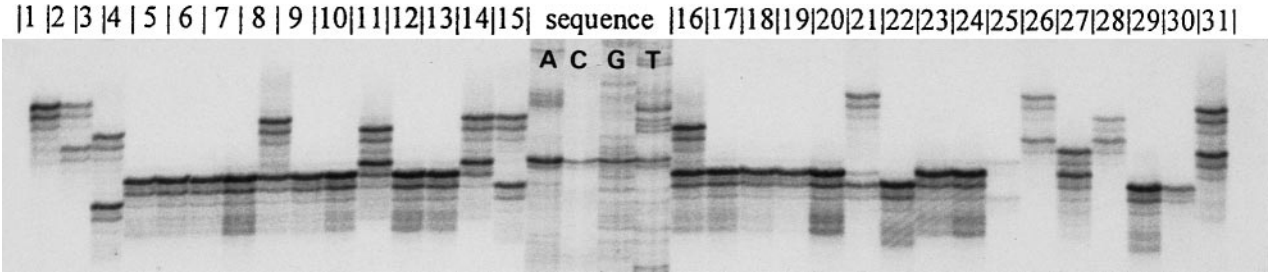

Fig. 1 An example of allelic polymorphism at microsatellite locus CNZ02 in 31 individuals of 14 coconut populations from across the geographic range. Numbers 1–31 refer to the sample identification numbers listed in Table 1. The sequence lanes are used to help determine the allele size in base pairs. In all cases, the darkest band was taken as the allele

Table 2 Sequences of the primer pairs defining the 39 microsatellites used in this study

1	CNZ01F ATGATGATCTCTGGTTAGGCT CNZ01R AAATGAGGGTTTGGAAGGATT	21	CNZ29F TAAATGGGTAAGTGTTTGTGC CNZ29R CTGTCCTATTTCCCTTTCATT
2	CNZ02F CTCTTCCCATCATATACCAGC CNZ02R ACTGGGGGGATCTTATCTCTG	22	CNZ31F GAAAGAATGTGGTGTTGACAT CNZ31R TAGTGCTAAGTCGATGGAAAT
3	CNZ03F CATCTTTCATCATTTAGCTCT CNZ03R AAACCAAAAGCAAGGAGAAGT	23	CNZ32F TTGATCCCTAAGAGAAGGATC CNZ32R GAAGAACACCAATGAGGTAAA
4	CNZ04F TATATGGGATGCTTTAGTGGA CNZ04R CAAATCGACAGACATCCTAAA	24	CNZ33F TTGCCCTATGACATTAAAAGA CNZ33R GAGGTCAAAGTTATTTCCGAT
5	CNZ05F CTTATCCAAATCGTCACAGAG CNZ05R AGGAGAAGCCAGGAAAGATTT	25	CNZ34F CATGTCGATTAATTATCCCAA CNZ34R TGCAAATATGAATGCAAACAC
6	CNZ06F ATACTCATCATCATACGACGC CNZ06R CTCCCACAAAATCATGTTATT	26	CNZ37F GTGGATAACTCATTTCAGGTG CNZ37R TAAGAAAGCTGAGAGGGAGAT
7	CNZ09F ATCTACCAGTGTGGTCCTCTC CNZ09R ACCAGGAAAAAGAGCGGAGAA	27	CNZ40F CTTGATTGCTATCTCAAATGG CNZ40R CTGAGACCAAATACCATGTGT
8	CNZ10F CCTATTGCACCTAAGCAATTA CNZ10R AATGATTTTCGAAGAGAGGTC	28	CNZ42F TGATACTCCTCTGTGATGCTT CNZ42R GTAGATTGTGGGAGAGGAATG
9	CNZ12F TAGCTTCCTGAGATAAGATGC CNZ12R GATCATGGAACGAAAACATTA	29	CNZ43F TCTTCATTTGATGAGAATGCT CNZ43R ACCGTATTCACCATTCTAACA
10	CNZ13F TATGCTATTCACTCATTTTTCG CNZ13R ACTCTGTTTCACGATCAAAAA	30	CNZ44F CATCAGTTCCACTCTCATTTC CNZ44R CAACAAAAGACATAGGTGGTC
11	CNZ16F TAAACTGAAAATAGCATGGGA CNZ16R CTGACGGAATCTGTAAATTTG	31	CNZ46F TTGGTTAGTATAGCCATGCAT CNZ46R AACCATTTGTAGTATACCCCC
12	CNZ17F ATGTAAAGAAAGTAGGGAGGC CNZ17R CATAGGTTATCATGCAGAGCT	32	CN11E6F TACTTAGGCAACGTTCCATTC CN11E6R TAACCAGAAAGCAAAAAGATT
13	CNZ18F ATGGTTCAGCCCTTAATAAAC CNZ18R GAACTTTGAAGCTCCCATCAT	31	CN1C6F AGTATGTGAGTAGGATTATGG CN1C6R TTCCCTGGACCCTTATCTCTT
14	CNZ19F GAAGAAGGTGCTGGTTCTGTC CNZ19R GGAGAAGAGAAGGAACAGAGA	34	CN2A4F CAGGATGGTTCAAGCCCTTAA CN2A4R GGTGGAAGAGGGAGAGATTGA
15	CNZ20F GTGGGACACATTTATCATTTA CNZ20R CAATTTTGCTGAAGAGGTAAA	35	CN2A5F AAGGTGAAATCTATGAACACA CN2A5R GGCAGTAACACATTACACATG
16	CNZ21F ATGTTTTAGCTTCACCATGAA CNZ21R TCAAGTTCAAGAAGACCTTTG	36	CN11E10F AGAGAGAGTAAATGGGTAAGT CN11E10R CCCTTTCATTTTTCCTTATTC
17	CNZ22F ACAAAGTCAAAGTTGATCAGC CNZ22R ATTACACACACACACACACCA	37	CN11A10F GTTGGAGATTTAATTTTCTTG CN11A10R CCCAATAATATTTTATAACAG
18	CNZ23F ATCAAAACATGACACCGTAAC CNZ23R CTGATAGATGACAAGGTGTGG	38	CN1G4F GTCGTCCTATACTCATCATCA CN1G4R GATGCGTATGAGATGTGAGAG
19	CNZ24F TCCTAAGCTCAATACTCACCA CNZ24R CGCATTGATAAATACAAGCTT	39	CN1H2F TTGATAGGAGAGCTTCATAAC CN1H2R ATCTTCTTTAATGCTCGGAGT
20	CNZ26F CTAGGCTCCCCATGTGTTTTT CNZ26R CACTGCTGTTGTACACCTCCA		

populations of the Pacific and of South East Asia, consistent with the hypothesis suggesting these regions to be the putative origins of coconut (Lebrun et al. 1998 a, b). Two of the four Dwarf populations, known to be self-pollinating, were homozygous at all 37 loci, whilst one had a single heterozygous locus. The fourth Dwarf, Niu Leka Dwarf, which is known to be cross-pollinating, was heterozygous at several loci. The East African coconuts had higher heterozygosities than those from West Africa, and the Tonga and Fijian populations had

Table 3 SSR repeat-type, number of alleles per locus, allele size ranges and gene diversity (D) for the SSRs tested on 2–3 samples of 14 coconut populations (31 individuals)

SSR	Repeat type	Allele number	Size range	D	SE(D)
CNZ01	$(CT)_{15}$ $(CA)_9$	7	109–131	*a	*a
CNZ02	$(GA)_{15}$	10	143–161	0.7133	0.0546
CNZ03	$(GA)_7$	3	91–97	0.4688	0.0623
CNZ04	$(CT)_{29}$ $TT(CA)_{10}$	11	118–174	0.8033	0.0358
CNZ05	$(CT)_{17}(GT)_7$	7	151–178	0.5765	0.0616
CNZ06	$(CT)_{15}$	7	71–93	0.8033	0.0173
CNZ09	$(CT)_{13}(CA)_6(CT)_{19}(CA)_8$	11	118–142	0.7820	0.0395
CNZ10	$(CT)_{18}(GT)_{17}$	13	105–175	0.8096	0.0328
CNZ12	$(CT)_{15}$	7	205–229	0.7237	0.0264
CNZ13	$(GCC)_6$	*b	*b	*b	*b
CNZ16	$(GTT)_6(GCT)_2(CTT)_8TTTGTGC(GT)_8CT(GT)_7CT(GT)_6$	5	174–185	0.6597	0.0333
CNZ17	$(CA)_{12}$ $(GA)_{11}$	8	89–137	0.6108	0.0644
CNZ18	$(CT)_{15}$ $TT(CT)_3$	13	103-127	0.8923	0.0132
CNZ19	$(CT)_{15}$ $(CA)_5CT(CA)_3(CT)_2(CA)_6$	9	171–193	0.8137	0.0259
CNZ20	$(CA)_{13}$	5	89–105	0.6660	0.0348
CNZ21	$(CT)_{30}$	14	231–261	0.9037	0.0100
CNZ22	$(GT)_{18}$	*c	*c	*c	*c
CNZ23	$(GA)_{18}$	5	142–164	0.6805	0.0285
CNZ24	$(CT)_{19}(GT)_{19}$	10	221–271	*a	*a
CNZ26	$(GT)_{17}(GA)_{19}$	12	206–262	0.8642	0.0158
CNZ29	$(GT)_{22}(GA)_2CA(GA)_{11}$	13	102–180	0.7773	0.0368
CNZ31	$(GT)_{14}$	5	181–193	0.7560	0.0201
CNZ32	$(GA)_{18}$	8	138–156	0.5375	0.0707
CNZ33	$(GT)_{11}$	2	152–169	0.4870	0.0202
CNZ34	$(CT)_{15}(GT)_7$	11	131–185	0.8424	0.0183
CNZ37	$(CT)_7CC(CT)_{13}(GT)_5N_{23}(GT)_{29}$	14	196–248	0.8845	0.0171
CNZ40	$(CT)_{20}$	8	132–160	0.8189	0.0193
CNZ42	$(CT)_{16}$	8	149–175	0.6873	0.0352
CNZ43	$(GA)_{21}$	10	175–217	0.8028	0.0291
CNZ44	$(GA)_{15}$	9	148–172	0.8215	0.0761
CNZ46	$(CT)_{24}$	9	97–117	0.8304	0.0163
CN1H2	$(GA)_{18}$	9	188–208	0.7680	0.0299
CN1C6	$(GT)_{11}$ $TT(GT)_5$	6	175–184	0.6993	0.0346
CN2A5	$(CT)_{30}$	12	150–190	0.7674	0.0457
CN1G4	$(CT)_{15}$	7	112–132	0.7854	0.0212
CN11E6	$(CT)_{21}$	16	85–128	0.7711	0.0511
CN2A4	$(CT)_{15}TT(CT)_3$	13	87–111	0.8392	0.0239
CN11E10	$(GT)_{22}(GA)_{14}$	9	99–151	0.7305	0.0401
CN11A10	$(CT)_{30}$	12	81–119	0.9027	0.0090

*a: D not estimated, as not all 14 populations could be analysed due to repeated amplification failure for some samples
*b: alleles were very close in size and, together with stutter problems, accurate reading was not possible
*c: complex multi-locus pattern that was not scorable

generally different alleles from the other South-Pacific populations.

A cluster analysis (UPGMA) was performed using the similarity matrix based on the proportion of shared alleles across the 37 SSR loci. The resultant dendrogram constructed from this analysis is shown in Fig. 2 a. In all cases, the two to three samples of the same population clustered together in the dendrogram. There was generally a good grouping of the Tall populations into different geographic regions (Fig. 2 a). The West African population, which is grouped with the Sri Lankan population, is thought to have been imported from this region. The populations from Mozambique and Andaman (an Indian archipelago, located north of Sumatra) share a large number of alleles with the previous group and are in the same cluster. The Tall populations from South East Asia, Papua New Guinea and the Solomon Islands form a broad cluster, including the Panamean population. Considering the Dwarfs, although collected from different regions, the South Pacific Kiribati population clustered together with those from Papua New Guinea and from Malaysia. This is in agreement with the available strong evidence suggesting a common origin of the Dwarfs (Lebrun et al. 1998 b). The Fiji (allogamous Dwarf) and Tonga (Tall) cultivars, also from the South Pacific, in contrast formed a quite separate group.

A PCO analysis was also carried out on the similarity matrix, with scores on the resultant first two components plotted pair-wise in each case. Data are shown for the first two coordinate axes, representing 15.8% and 9.4% of the total variation, respectively (Fig. 2 b). The results showed similar groupings to the UPGMA analysis, although the clustering was not tight. One cluster comprised the Dwarf Kiribati, Papua New Guinea and Malaysian populations. Two other small clusters contained the West African-Sri Lankan and the East African-Anda-

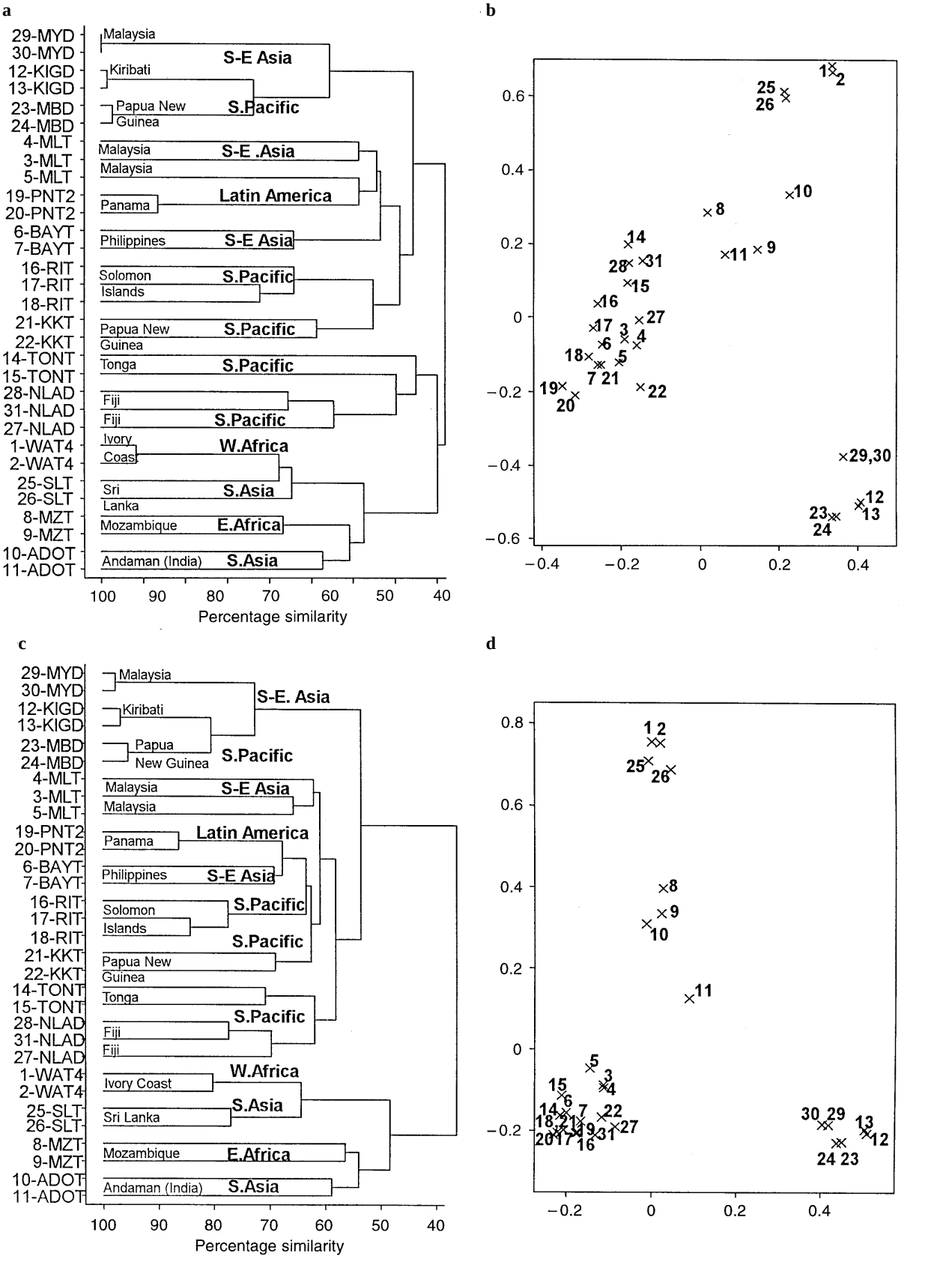

Fig. 2 Dendrograms and PCO plots from SSR (*a, b*, respectively) and AFLP (*c, d*, respectively) data for the 31 individuals from 14 coconut populations. The AFLP results shown are for the polymorphic bands (303) from 12 primer combinations. The dendrograms are based on the UPGMA method for cluster analysis of the Jaccard coefficient and the PCO plots are of components 1 (vertical axis) vs 2 (horizontal axis). Full details of the population codes given on the y axes of the dendrograms and the numerical codes given on the PCO plots are listed in Table 1

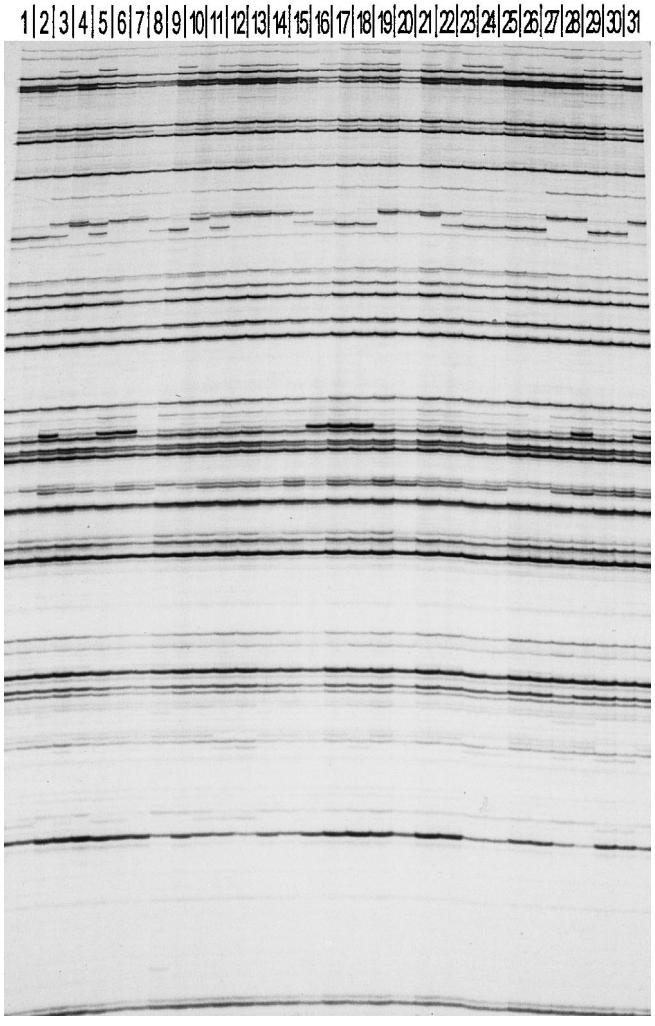

Fig. 3 AFLP profiles for the 31 coconut individuals, from the primer combinations *Mse*I+GAA and *Pst*I+AAC. The code numbers *1–31*, refer to the sample identification numbers listed in Table 1

man groups respectively. The largest group was formed mainly by the remaining South Pacific and South East Asian populations. These resolved into South Pacific and South East Asian groups when the second and third coordinates of the PCO were plotted (data not shown).

AFLP analysis with 12 primer combinations on the 14 populations yielded a total of 1106 bands, of which 303 (27%) were polymorphic. An example is shown in Fig. 3. Three primer combinations (*Mse*I+GAA and *Pst*I+AAA; *Mse*I+GAA and *Pst*I+CC; *Mse*I+ACA and *Pst*I+AAA) gave over 100 bands per gel, with an average of 25% polymorphic bands. Seven primer combinations (*Mse*I+GAA and *Pst*I+AAC; *Mse*I+ACA and *Pst*I+AAC; *Mse*I+ACA and *Pst*I+AAG; *Mse*I+GAA and *Pst*I+CAG; *Mse*I+GAA and *Pst*I+CAT; *Mse*I+ACA and *Pst*I+CAG; *Mse*I+ACA and *Pst*I+CAT) yielded between 80 and 100 bands per gel, and gave an average of 30% polymorphic bands. Two primer combinations (*Mse*I+ACA and *Pst*I+CAA; *Mse*I+ACA and *Pst*I+CAC) gave 71 and 60 bands, of which 37% and 13% were polymorphic, respectively.

A Jaccard similarity matrix from the combined data set of all 12 primer combinations, was constructed and analysed by UPGMA and PCO analysis. The dendrogram derived from the UPGMA analysis (Fig. 2 c) showed similar relationships among populations as shown by the SSR data, with the exception that the Tonga and Fiji populations cluster together with the South East Asian and South Pacific Talls.

The PCO plot from the AFLP data, in which 21.7% and 11.9% of the variation was accounted for in the first and second components respectively (Fig. 2 d), resolved tighter clusters than that from the SSR data (Fig. 2 b). One cluster comprised the three autogamous Dwarf populations. Populations of the Ivory Coast and Sri Lanka formed another cluster, whilst the Mozambique samples were grouped somewhat separately with the Andaman population. The remaining populations formed another group. The slightly higher resolution of genetic similarities by AFLPs, compared with SSRs, found here is consistent with previous reports in other crops where it has been suggested that the high polymorphism of SSRs may render them less suitable for determining relationships among cultivars (e.g. Powell et al. 1996 b; Russell et al. 1997).

The phenetic relationships among coconut populations revealed by SSR and AFLP analyses are in concordance with previous results obtained using 40 polymorphic bands from nine RFLP heterologous probes (Lebrun et al. 1998 a, b). Using RFLPs, two main groups were identified comprising populations from the Indian Ocean and Africa (including WAT4 and SLT analysed here) and from the Far East and South Pacific (including RIT, NLAD, MLT, MYD analysed here), respectively. A third group, including the populations from the Mozambique and the Andamans, showed characteristics of each of the preceding groups. This is probably due to the intercrossing of populations of both origins (due to transport by Man, at least in the case of East Africa) and explains the high polymorphism found in these populations, compared to Sri Lanka and West Africa. The RFLPs showed that the West African populations were similar to the Sri Lankan and Indian types, as found in this study. This grouping adds weight to the hypothesis that movement of nuts, originating from this region of India and Sri Lanka, have extended the range of coconut along the Atlantic coasts of Africa, probably at the end of the XVth century. Similarly, the populations from Panama were related to South East Asian and South Pacific types, again as found here. Coconut was discovered by the Spaniards on the Pacific coast of Panama (Zizumbo-Villareal 1996), although it is not considered to be indigenous (Harries 1978). Considering the distances, transport by tidal action is unlikely. On the other hand, there is evidence that America was occasionally visited by travellers coming either from Polynesia or from South East Asia (Langdon 1993). The main differences between the two studies are that SSRs separate closely related cultivars more accurately, while the differences between

large groups of populations (South East Asia and Pacific on one side and South Asia and West Africa on the other side) are more obvious with RFLPs.

Our data show that both AFLPs and SSRs are informative in evaluating genetic diversity in coconut. AFLPs resolved relationships among populations with a better resolution than SSRs, but SSRs were more informative in revealing additional data on differences in allelic combinations between and within populations and highlighting regions where allelic compositions were unusual or more diverse. Approximately 3100 PCR amplifications and 79 sequencing gels were performed to complete the analysis. Of these, about 600 PCR and 12 gels were used for the AFLP work and 2500 PCR and 67 gels for the microsatellite work. However, the high levels of polymorphism revealed by the SSRs indicate that populations could be identified and information on allelic diversity obtained with relatively few loci. Together with multiplexing of the SSRs chosen, the number of gels required for screening could, thus, be significantly reduced. The SSRs would be easily exchangeable between different groups and the data are more amenable for use in databases. For gene-banking purposes, given the amount of additional information obtained and the exchangeability of the data, we consider SSRs to provide the most informative means for evaluating genetic diversity in coconut populations.

Acknowledgements We gratefully acknowledge the financial support of a grant from Burotrop. We thank Marc Delorme Research Station (CNRA, Côte d'Ivoire), for providing the samples and also Mr. Roger Smith of Burotrop, Dr. Pons Batuga of COGENT and Dr. Toby Hodgkin of IPGRI (International Plant Genetic Resources Institute) for their interest and encouragement. IACR receives grant-aided support from the Biotechnology and Biological Sciences Research Council of the United Kingdom.

References

Akpan EEJ (1994) Evaluation of Tall coconut (*Cocos nucifera* L) genotypes within the Nigerian coconut germplasm bank. Oleagineux 49:13–20

Ashburner GR, Thompson WK, Halloran GM, Foale MA (1997 a) Fruit component analysis of south Pacific coconut palm populations. Gene Resour Crop Evol 44:327–335

Ashburner GR, Thompson WK, Halloran GM (1997 b) RAPD analysis of South Pacific coconut palm populations. Crop Sci 37:992–997

Carpio CB (1982) Biochemical studies of *Cocos nucifera* L. Kalikasan. Philippine J Biol 11:319–338

Fernando WMU, Perera L, Peries RRA (1997) An overview of breeding research in coconut – the Sri Lankan experience. Outlook Agric 26:191–198

Genstat 5 Committee (1993) Genstat 5 Reference Manual, Release 3. Oxford University Press.

Gupta PK, Balayan HS, Sharma PC, Ramesh B (1996) Microsatellites in plants: a new class of molecular markers. Curr Sci 70:45–54

Harries HC (1978) The evolution dissemination and classification of *Cocos nucifera* L. Bot Rev 44:205–317

Harries HC (1990) Malesian origin for a domestic *Cocos nucifera*. In: Baas P et al. (ed) The plant diversity of Malesia. Kluwer, Dordrecht, the Netherlands, pp 351–357

Langdon R (1993) The banana as a key to early American and Polynesian history. J Pacific Hist 28:15–35

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

Lebrun P, Grivet L, Baudouin L (1998 b) Dissémination et domestication du cocotier à la lumière des marqueurs RFLP Plantations. Recherch, Dévelop 5:233–245

Nei M (1973) Analysis of gene diversity in subdivided populations. Proc Natl Acad Sci USA 70:3321–3323

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 Genet 96:545–550

Powell W, Machray GC, Provan J (1996 a) Polymorphism revealed by simple sequence repeats. Trends Plant Sci 1:215–222

Powell W, Morgante M, Andre C, Hanafey M, Vogel J, Tingey S, Rafalski A (1996 b) The comparison of RFLP, RAPD, AFLP and SSR (microsatellite) markers for germplasm analysis. Mol Breed 2:225–238

Rhode W, Kullaya A, Rodriguez J, Ritter E (1995) Genome analysis of *Cocos nucifera* L. by PCR amplification of spacer sequences separating a subset of copia-like *Eco*RI repetitive elements. J Genet Breed 49:179–186

Rivera R, Edwards KJ, Barker JHA, Arnold GM, Ayad G, Hodgkin T, Karp A (1999) Isolation and characterisation of polymorphic microsatellites in *Cocos nucifera* L. Genome 42:1–8

Russell JR, Fuller JD, Macaulay M, Hatz BG, Jahoor A, Powell W, Waugh R (1997) Direct comparisons of levels of genetic variation among barley accessions detected by RFLPs, AFLPs, SSRs and RAPDs. Theor Appl Genet 95:714–722

Sugimura Y, Itano M, Salud CD, Otsuji K, Yamaguchi H (1997) Biometric analysis on diversity of coconut palm: cultivar classification by botanical and agronomical traits. Euphytica 98:29–35

Vos P, Hogers R, Bleeker M, Rijans M, Van de Lee T, Hornes M, Frijters A, Pot J, Peleman J, Kuiper M, Zabeau M (1995) AFLP: a new technique for DNA fingerprinting. Nucleic Acids Res 23:4407–4414

Whitehead RA (1976) Coconut. In: Simonds NW (ed) Evolution of crop plants. Longman, London, pp 221–225

Zabeau M, Vos P (1993) Selective restriction fragment amplification: a general method for DNA fingerprinting. European Patent Application, publication no. EP 0534858-A1, No. 92402629.7

Zizumbo-Villarreal D (1996) History of coconut (*Cocos nucifera* L.) in Mexico: 1539–1810. Gene Resour Crop Evol 43:505–515