Genet Resour Crop Evol (2010) 57:203–217
DOI 10.1007/s10722-009-9463-x

RESEARCH ARTICLE

Ambiguous genetic relationships among coconut (*Cocos nucifera* L.) cultivars: the effects of outcrossing, sample source and size, and method of analysis

Margarita Mauro-Herrera · Alan W. Meerow · Lalith Perera · Joanne Russell · Raymond J. Schnell

Received: 17 March 2009 / Accepted: 23 June 2009 / Published online: 31 July 2009

Abstract A prior analysis of eight coconut cultivars with 15 microsatellite (SSR) markers drew unexpected relationships between two of the out-crossing tall cultivars evaluated: 'Atlantic Tall' and 'Panama Tall'. We further investigated the relationships between these eight cultivars by increasing the number of individuals studied (particularly for 'Atlantic Tall' and 'Panama Tall'), by including 28 more molecular markers, and by adding two other cultivars to our analysis. Our results show that five to ten coconut individuals do not represent a dependable sample to withdraw conclusions regarding cultivar/variety relationships, particularly when studying out-crossing genotypes. As suggested in the prior study, a high level of hybridization was observed between the 'Atlantic Tall' and 'Panama Tall' cultivars. However, at this time we were able to identify distinct groups for each one of these two cultivars. The two clustering methods used (Neighbor Joining, NJ and Unweighted Pair Group Method with Arithmetic mean, UPGMA) produced dendrograms that resolved contrasting cultivar relationships, especially for the 'Atlantic Tall' and 'Panama Tall' cultivars. We discuss the implications of our results in regard to current scenarios of coconut domestication and future considerations when assessing genetic relationships among different varieties.

Keywords 'Atlantic Tall' · *Cocos nucifera* · Coconut · Clustering method · Microsatellite DNA · 'Panama Tall' · Population structure · Sample size · WRKY

M. Mauro-Herrera · A. W. Meerow (✉) · R. J. Schnell
USDA-ARS-SHRS, National Germplasm Repository, 13601 Old Cutler Rd., Miami, FL 33158, USA
e-mail: alan.meerow@ars.usda.gov

L. Perera
Genetics and Plant Breeding Division, Coconut Research Institute, Lunuwila 61150, Sri Lanka

J. Russell
Scottish Crops Research Institute, Invergowrie, Dundee DD2 5DA, UK

Introduction

The high levels of genetic diversity in coconut, *Cocos nucifera* L., are the result of natural evolution and adaptation, as well as human involvement in the exploitation of the species (Harries 1978, 2001; Harries et al. 2004). Coconut grows in the tropical environments of Africa, America, Asia and Oceania, and is an important export crop for oil and fiber in many tropical countries. For millennia, humankind has found myriad uses for almost all parts of the coconut palm, and this history has played an important role in shaping the phenotypic diversity of this cultigenic species (Harries 2001). Coconut cultivars are generally classified into the Tall and Dwarf types. The tall type is primarily

out-crossing while the dwarf type is mainly selfing (with a few exceptions).

A number of marker types have been used in the study of coconut (e.g., Ashburner et al. 1997; Lebrun et al. 1998; Perera et al. 1998, 1999, 2000; Rivera et al. 1999; Rohde et al. 1995; Teulat et al. 2000). Based on the results of these studies, a classification was proposed by Lebrun et al. (2005). The classification proposed two major cultivar groups: the Pacific group with five sub-groups (Southeast Asia, Melanesia, Micronesia, Polynesia and the Pacific coast of Central and South America), and the Indo-Atlantic group (Lebrun et al. 2005).

A population structure and genetic diversity analyses of a cultigenic species is a usual first common task in the understanding of intra-specific relationships (Van den Berg and Jacobs 2007). Assuming adequate sampling, both of markers and individuals, the analysis should resolve the population structure of the species, and the levels of diversity within each genealogy. The analysis of the relationships among genetic groups may also convey information regarding their origins and diversification (Van den Berg and Jacobs 2007). Several factors should be considered while conducting these studies as they affect the accuracy of the conclusions. In addition to adequate sample size, sample source is of concern. Sample source is particularly important for samples obtained from different collections of the presumed identical cultivar (same name or similar phenotypes), which sometimes correspond to different genotypes (e. g., 'San Ramon' coconut cultivars from the Philippines and Sri Lanka, Perera et al. 2003). Lastly, the level of out-crossing (if present in the species) is another factor to consider while carrying out the study, particularly if the samples originate from a diverse germplasm collection.

We previously evaluated several coconut cultivars commonly used in the landscape of Florida, USA in order to determine their population structure and genetic relationships (Mauro-Herrera et al. 2007; Meerow et al. 2003). The results of those studies were perplexing as two of the cultivars evaluated (i.e. 'Atlantic Tall' and 'Panama Tall'), which have been classified as belonging, respectively, to the two major cultivar groups (Lebrun et al. 2005), were grouped together. The discrepancy of our results with those of other studies (reviewed by Lebrun et al. 2005) led to the further investigation of the plant material/cultivars analyzed. In this paper, we address the possible factors that might have affected our previous results. The same genotypes studied in Meerow et al. (2003) are used but the number of individuals sampled on a per cultivar basis is increased, two additional cultivars are sampled, and the number of markers is increased to 41. We discuss the implications of our results in regards to the origin of these two cultivars.

Materials and methods

Plant material

Two hundred and eighteen genotypes consisting of ten coconut cultivars, two off-types, a 'Maypan' (a hybrid between a 'Malayan Dwarf' and a 'Panama Tall') and an undetermined tall were evaluated in this study. One hundred and ten of these genotypes were previously studied with 15 microsatellite markers (Meerow et al. 2003). The description and provenance of the plant material is presented in Table 1.

Data collection

Genomic DNA samples were collected and extracted as in Meerow et al. (2003). Polymerase chain amplifications and data collection were performed as in Mauro-Herrera et al. (2006). A total of 41 codominant markers were analyzed: 28 microsatellite or simple sequence repeat (SSR) markers and 13 WRKY-derived markers (Mauro-Herrera et al. 2006, Table 2). Twenty of the 28 SSR markers (labeled CAC*n*, Table 2) were developed by Perera et al. (1999, 2003 and seven new markers, Table 2). The other eight SSR markers (labeled CN1G4, CN1H2, CN2A5 and CNZ*n*, Table 2) were developed by Rivera et al. (1999). The annealing temperatures for our PCR conditions for the SSR primers were determined via a gradient and are reported on Table 2. Only 13 of the 15 SSR markers analyzed by Meerow et al. (2003) were used in this study since the other two (CAC8 and CAC20) had problems as determined by the parentage analysis performed in that study. The primer sequences and amplification conditions for the 13 WRKY-derived markers were reported in Mauro-Herrera et al. (2006) and in Mauro-Herrera et al. (2007). Twelve of the 13 WRKY-derived markers contained single nucleotide polymorphisms (SNPs) and were analyzed via single-strand conformation polymorphism (SSCP). The

Table 1 Designation, sample size, individual labeling and seed source for the 218 coconut (*Cocos nucifera* L.) genotypes studied

Cultivar	Sample size	Labeling of individuals	Source
'Atlantic Tall'	9	JT01–JT09	Fort Lauderdale, FL. Local, open pollinated
'Atlantic Tall'	12	AT01–AT12	Caribbean coast of Costa Rica.
'Atlantic Tall'	7	JTS01–JTS07	Palm Beach, FL. Local, open pollinated
'Chowghat'	7	CH01–CH07	Fort Lauderdale, FL. Second generation seedlings grown from seed sent from the Coconut Industry Board (CIB), Jamaica
'Fiji Dwarf' ('Niu Leka')	6	UF14–UF18 and FDtrue	Coconut Industry Board (CIB), Jamaica
'Fiji Dwarf' ('Niu Leka')	48	FD01–FD33 and FDp1–FDp15	Open pollination at Fort Lauderdale, FL
'Hawaiian Tall'	2	HT01, HT02	Hawaii, maintained in the SHRS-Miami, FL
'Hawaiian Tall'	11	HT03–HT013	Open pollination at UF-Fort Lauderdale, FL Several half sibs from a single maternal plant
'Green Malayan Dwarf'	10	GrMD01–GrMD10	Coconut Industry Board (CIB), Jamaica
Off-type 'Green Malayan Dwarf'	1	otGrMD	Coconut Industry Board (CIB), Jamaica
'Red Malayan Dwarf'	10	RMD01–RMD10	Coconut Industry Board (CIB), Jamaica
'Red Malayan Dwarf'	5	RMD011–RMD15	Miami, FL. Local source
'Red Malayan Dwarf'	4	RMD16–RMD19	Coconut Industry Board (CIB), Jamaica
Off-type 'Red Malayan Dwarf'	1	otRMD	Coconut Industry Board (CIB), Jamaica
'Yellow Malayan Dwarf'	5	YMD01–YMD05	Miami, FL. Local source
'Green Niño'	25	N01–N25	Coconut Industry Board (CIB), Jamaica
'Maypan'	1	M01	Coconut Industry Board (CIB), Jamaica
'Panama Tall'	6	PT01–PT06	Coconut Industry Board (CIB), Jamaica
'Panama Tall'	10	PT1a–PT10a	Miami, FL. Local source
'Panama Tall'	11	PTCR1–PTCR11	Pacific coast of Costa Rica
'Red Spicata'	25	RS01–RS25	Miami, FL. Local source
'Red Spicata'	1	RS26	Coconut Industry Board (CIB), Jamaica
Undetermined tall	1	Tallund	Unknown

remaining WRKY marker contained an SSR. Allele detection and analysis for both SSR and WRKY markers were done as in Mauro-Herrera et al. (2006).

Data analysis

Several of the formatted files for the programs that we used were produced with Convert (Glaubitz 2004). The population genetic structure analyses were performed with the Bayesian model-based clustering program STRUCTURE (Pritchard et al. 2000; Falush et al. 2003). A first population structure analysis was done with all 218 genotypes in order to determine the number of genetic groups and hybrids. We used this program to assign individuals to genetic groups since it does not require a priori information regarding individual provenance, and it proved useful at identifying known hybrids/off-types in a preliminary study (Mauro-Herrera et al. 2007). This feature was important since we suspected the presence of hybrids in two of the cultivars studied. The number of populations/clusters (k) was determined following the Δk method of Evanno et al. (2005). The admixture model with correlated allele frequencies was used with every run. To determine k, a total of 20 runs for each k ranging from four to eighteen using 10^4 iterations for the burn-in period and 2×10^4 iterations for the running period were used, and the k value with highest Δk was determined. The STRUCTURE analysis with the determined k value was performed with 10^5 iterations for the burn-in period and 10^6 iterations for the running period. The results of this first analysis were used to remove every probable hybrid genotype (a total of 86) from a second STRUCTURE analysis performed. K was determined and the final STRUCTURE analysis completed as before with the reduced file (132 individuals). The graphic displays of the

Table 2 Information for the 41 molecular markers used in the analysis of the 218 coconut genotypes

Locus	Reference for previously published or primer sequences and GenBank acc. nos. for microsatellite markers published in this paper	Annealing temperature (°C)	Number of alleles	PIC	Gene diversity
CAC4	Perera et al. (1998, 1999)	58	7	0.674	0.412 ± 0.26
CAC6	Perera et al. (1998, 1999)	58	9	0.741	0.388 ± 0.36
CAC11	Perera et al. (1998, 1999)	58	3	0.375	0.201 ± 0.18
CAC13	Perera et al. (1998, 1999)	58	2	0.363	0.203 ± 0.18
CAC21	Perera et al. (1998, 1999)	58	2	0.373	0.281 ± 0.18
CAC23	Perera et al. (1998, 1999)	58	3	0.584	0.290 ± 0.25
CAC27[a]	F:TCCAAGTCCTTGGTTAGC R:AAAACACAACCCACCTAAG FJ499399	56	3	0.376	0.257 ± 0.16
CAC46[a]	F:GATGGTTGGATATCATTCTTG R:TTGACCTATCAAATGTGCC FJ499400	52	13	0.796	0.537 ± 0.22
CAC52	Perera et al. (1998, 1999)	58	8	0.699	0.409 ± 0.30
CAC56	Perera et al. (1998, 1999)	58	8	0.734	0.437 ± 0.22
CAC65	Perera et al. (1998, 1999)	58	7	0.666	0.416 ± 0.25
CAC67[a]	F:GGAGAAACGGTATACCAGAG R:CCTCATTTAGATGCCCTATC FJ499401	56	7	0.741	0.459 ± 0.31
CAC68	Perera et al. (1998, 1999)	58	5	0.534	0.312 ± 0.29
CAC69[a]	F:TATAAATGGGTAGCCCTGAG R:TGAATAGGTTGGTGAATGTG FJ499402	58	5	0.555	0.354 ± 0.21
CAC70[a]	F:AACAAATGAAAACTTGATTCC R:AACTTGCCATGTTTTACTTGT FJ499403	52	6	0.605	0.378 ± 0.26
CAC71	Perera et al. (1998, 1999)	58	5	0.462	0.362 ± 0.22
CAC72	Perera et al. (1998, 1999)	58	5	0.547	0.343 ± 0.25
CAC75[a]	F:GTTTCACCTTGTACTCTGTCC R:GAGAAATGGAAAACTTTTGTG FJ499404	56	10	0.816	0.418 ± 0.21
CAC77[a]	F:CAGAGGTCACAACCATATTG R:CTTTAGCTATTTGTTCCAAGG FJ499405	56	11	0.677	0.377 ± 0.36
CAC84	Perera et al. (1998, 1999)	58	3	0.301	0.290 ± 0.23
CN1G4	Rivera et al. (1999)	58	9	0.776	0.435 ± 0.20
CN1H2	Rivera et al. (1999)	56	6	0.713	0.344 ± 0.26
CN2A5	Rivera et al. (1999)	54	14	0.822	0.491 ± 0.28
CNZ02	Rivera et al. (1999)	52	10	0.755	0.546 ± 0.26
CNZ05	Rivera et al. (1999)	52	10	0.523	0.428 ± 0.29
CNZ26	Rivera et al. (1999)	56	14	0.822	0.542 ± 0.25
CNZ29	Rivera et al. (1999)	58	10	0.838	0.429 ± 0.21
CNZ46	Rivera et al. (1999)	56	10	0.739	0.437 ± 0.27
CnWRKY-01	Mauro-Herrera et al. (2006)	54	4	0.400	0.108 ± 0.21

Table 2 continued

Locus	Reference for previously published or primer sequences and GenBank acc. nos. for microsatellite markers published in this paper	Annealing temperature (°C)	Number of alleles	PIC	Gene diversity
CnWRKY-02	Mauro-Herrera et al. (2006)	50	2	0.091	0.076 ± 0.13
CnWRKY-03	Mauro-Herrera et al. (2006)	55	5	0.497	0.370 ± 0.26
CnWRKY-04	Mauro-Herrera et al. (2006)	52	3	0.271	0.257 ± 0.19
CnWRKY-05	Mauro-Herrera et al. (2006)	50	3	0.411	0.289 ± 0.20
CnWRKY-06	Mauro-Herrera et al. (2006)	52	2	0.191	0.104 ± 0.17
CnWRKY-09	Mauro-Herrera et al. (2006)	52	2	0.332	0.317 ± 0.18
CnWRKY-10	Mauro-Herrera et al. (2006)	57	4	0.356	0.155 ± 0.20
CnWRKY-13	Mauro-Herrera et al. (2006)	50	4	0.331	0.191 ± 0.21
CnWRKY-14	Mauro-Herrera et al. (2006)	54	3	0.462	0.238 ± 0.27
CnWRKY-16	Mauro-Herrera et al. (2006)	50	3	0.269	0.162 ± 0.24
CnWRKY-19	Mauro-Herrera et al. (2006)	58	2	0.161	0.035 ± 0.11
CnWRKY-21	Mauro-Herrera et al. (2006)	58	3	0.396	0.316 ± 0.24

[a] Not previously published

STRUCTURE results were obtained with the program Distruct (Rosenberg 2004).

The reduced set of genotypes was used for all subsequent analyses. The descriptive statistics (allele number, average sample size, proportion of polymorphic loci, expected and observed heterozygosity and fixation index; Tables 2 and 3) were estimated with GDA 1.1 (Lewis and Zaykin 2001). Polymorphism information content (PIC) was estimated using Botstein's et al. (1980) formula implemented in PowerMarker (Liu and Muse 2005). Gene diversity and allelic richness were calculated with FSTAT (Goudet 2002). Gene diversity was estimated using Nei's unbiased estimator (Nei 1987). Allelic richness, an estimate of allele number independent of sample size (El Mousadik and Petit 1996; Petit et al. 1998), was calculated as

$$R_S = \sum_{i=1}^{n_u} \left[1 - \frac{\binom{2N - N_i}{2n}}{\binom{2N}{2n}} \right],$$

where $2N$ represents the number of sampled genes, $2n$ is a sub-sample of the sampled genes ($N \geq n$), and N_i is the number of type i alleles among the $2N$ genes. Genetic distances were estimated using the Cavalli-Sforza and Edwards' (1967) chord distance (Dc), and the Nei et al.'s (1983) Da genetic distance. Both of these estimates were the best at revealing correct tree topologies when compared with other genetic distances (Takezaki and Nei 1996). For cluster analysis we used POPULATIONS (Langella 2002) with both the neighbor-joining method (NJ, Saitou and Nei 1987) and the unweighted pair group method with arithmetic mean (UPGMA, Sneath and Sokal 1973). Statistical confidence of the trees was established via bootstrap analysis with 2,000 replications. We also tested how well each clustering method fit the genetic distance data using the program TreeFit (Kalinowski 2007). TreeFit calculates the proportion of variation (R^2) in the distance matrix that is captured by the tree topology. The closer R^2 is to 1.0, the better that tree represents a good summary of the genetic relationships shown in the distance matrix. Principal coordinate analysis (PCA) was performed with Nei et al.'s (1983) Da genetic distance estimates between individuals with the Multivariate Statistical Package (MVSP, Kovach Computing Services, Anglesey, Wales). PCA plots were generated with the SAS system ver. 9.0 (SAS Institute Inc., Cary, NC, USA).

Results

The initial STRUCTURE analysis with all 218 genotypes representing ten coconut cultivars, two off-types, a hybrid and an undetermined tall

Table 3 Genetic information for the ten cultivars studied and the two subgroups identified for two of them with STRUCTURE

Cultivar	Sample size[a]	Average allelic richness[b]	P[c]	He[d]	Ho[e]	f[f]	Gene diversity[g]
'Atlantic Tall'	13.5(28)	1.90 ± 0.58	0.93	0.434	0.312	0.287	0.439 ± 0.251
'Chowghat'	5.6(7)	1.01 ± 0.06	0.02	0.005	0.005	0	0.005 ± 0.031
'Fiji Dwarf' ('Niu Leka')-1	13.9(54)[h]	1.87 ± 0.55	0.85	0.421	0.318	0.251	0.425 ± 0.249
'Fiji Dwarf' ('Niu Leka')-2	11(54)[h]	1.59 ± 0.43	0.71	0.333	0.383	−0.160	0.330 ± 0.243
'Hawaiian Tall'	5(13)	1.47 ± 0.47	0.51	0.263	0.288	−0.105	0.260 ± 0.261
'Green Malayan Dwarf'	8.8(10)	1.40 ± 0.37	0.63	0.214	0.082	0.648	0.223 ± 0.208
'Red Malayan Dwarf'	18.4(19)	1.03 ± 0.11	0.17	0.017	0.011	0.365	0.017 ± 0.061
'Yellow Malayan Dwarf'	4.9(5)	1.47 ± 0.43	0.63	0.238	0.197	0.190	0.244 ± 0.224
'Green Niño'	10.5(25)	1.13 ± 0.25	0.24	0.068	0.036	0.472	0.069 ± 0.142
'Panama Tall'	6.8(16)	1.57 ± 0.49	0.71	0.290	0.155	0.486	0.303 ± 0.253
'Panama Tall'-Costa Rica	10.6(11)	1.48 ± 0.46	0.59	0.260	0.282	−0.089	0.259 ± 0.244
'Red Spicata'	16.4(26)	1.06 ± 0.10	0.27	0.030	0.031	−0.032	0.030 ± 0.051

These estimates were obtained with 132 individuals as all putative hybrids were removed from this analysis

[a] Average sample size over all loci, the number in the parentheses is the initial sample size for the corresponding cultivar

[b] Average allelic richness across loci, as a measure of the number of alleles independent of sample size

[c] P Proportion of polymorphic loci

[d] He Expected heterozygosity

[e] Ho Observed heterozygosity

[f] Fixation index

[g] Gene diversity, Nei's unbiased estimator (Nei 1987)

[h] The initial sample size for the 'Fiji Dwarf' cultivar includes both groups since the two genetic groups were established with the STRUCTURE analysis

established the presence of nine genetic groups: one for the 'Atlantic Talls', a single genetic group for the 'Chowghat', 'Red Malayan Dwarf' and 'Green Niño' cultivars, two groups in the 'Fiji Dwarf' ('Niu Leka') cultivar, one for the 'Hawaiian Talls', a single group for the 'Green' and 'Yellow Malayan Dwarf' cultivars, two groups for the 'Panama Tall' samples and one group for the 'Red Spicata' genotypes (Fig. 1A). STRUCTURE also identified the mixed makeup of the known hybrid (Maypan) and two off-types (Fig. 1a). Despite the fact that the STRUCTURE analysis identified a single genetic group for the 'Chowghat', 'Green Niño' and the 'Malayan Dwarf' cultivars, we chose to keep them separate for further analysis. The genetic distance estimates and the PCA analysis confirmed the proximity of these cultivars but also their distinctiveness (Table 4, Figs. 2, 3). In addition, substantial levels of hybridization/admixture were also detected for several of the cultivars studied: 'Atlantic Tall', 'Fiji Dwarf', 'Hawaiian Tall', 'Green Niño', 'Panama Tall', and for some of the 'Red Spicata' genotypes.

All nine 'Atlantic Tall' individuals used by Meerow et al. (2003) showed variable levels of admixture (JT01–JT09, first nine individuals from left in Fig. 1a). Most of them looked like the product of out-crossing with 'Green Niño', 'Green', 'Red' or 'Yellow Malayan Dwarf' genotypes while JT03 appeared as a mislabeled 'Panama Tall' genotype. These genotypes were obtained from Fort Lauderdale, FL, and represented progeny from an unknown number of 'Atlantic Tall' individuals that survived the LY epidemic of the late 1970s.

In regard to the twelve 'Atlantic Tall' genotypes collected from a pure grove of 'Atlantic Talls' on the Caribbean coast from Costa Rica (AT01–AT12, Fig. 1a), four showed some level of admixture with the 'Green Niño' and/or the 'Malayan Dwarf' cultivars, six others showed variable levels of admixture only with the 'Panama Tall' cultivar while two were completely homogeneous (true to type, all brown in Fig. 1a). Lastly, the seven 'Atlantic Tall' genotypes (JTS01–JTS07), which are seedlings from a surviving single mother tree located in West Palm

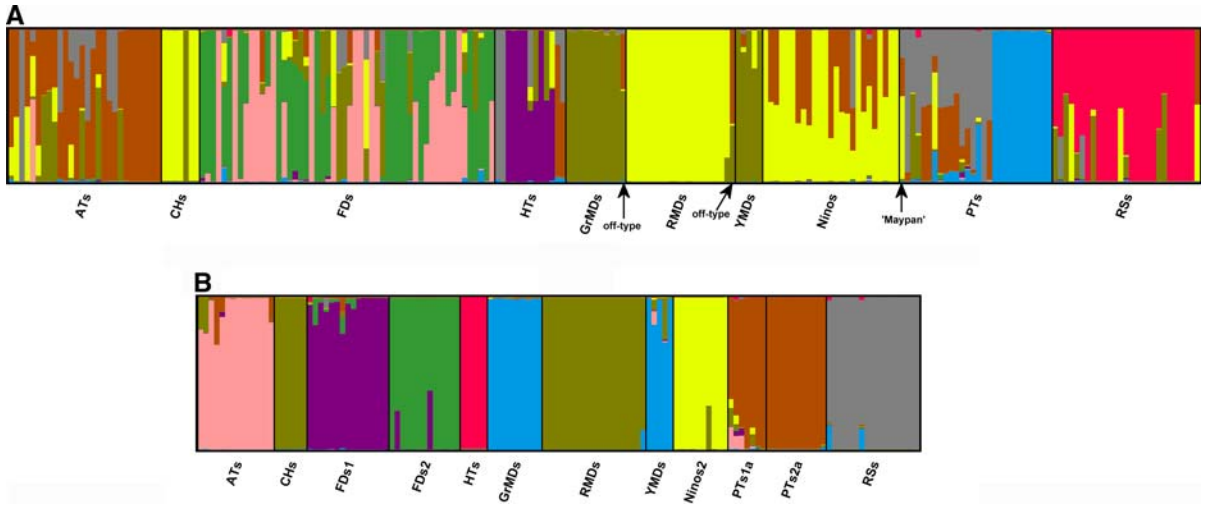

Fig. 1 Population structure results of coconut individuals representing ten cultivars analyzed with 41 molecular markers and the program STRUCTURE (Pritchard et al. 2000) and visualized with DISTRUCT (Rosenberg 2004). ATs: 'Atlantic Tall', CHs: 'Chowghat', FDs: 'Fiji Dwarf' ('Niu Leka'), HTs: 'Hawaiian Tall', GrMDs: 'Green Malayan Dwarf', RMDs: 'Red Malayan Dwarf', YMDs: 'Yellow Malayan Dwarf', Ninos: 'Green Niño', PTs: 'Panama Tall', RSs: 'Red Spicata'. A. Population structure results for all 218 coconut genotypes representing ten cultivars, two off-type Malayan Dwarf individuals, and one 'Maypan'. Individuals in Fig. **a** are represented in the same order as they are described in Table 1; **b** population structure results for 132 individuals representing all ten coconut cultivars after removing putative hybrid genotypes (including off-types). Based on the first analysis (Fig. 1a), the 'Fiji Dwarfs' were divided in two subgroups: FDs1 and FDs2, and the two subgroups for the 'Panama Talls', i.e. PTs1a and PTs2a, resolved as a single genetic group

Table 4 Nei et al.'s Da (1983) genetic distance (above diagonal) and Cavalli-Sforza and Edwards Dc (1967) genetic distance (below diagonal) between the ten coconut cultivars studied and the two genetic groups identified for each the 'Fiji Dwarf' and the 'Panama Tall' cultivars

	AT	CH	FD1	FD2	HT	GrMD	RMD	YMD	Niño	PT	PTCR	RS
AT		0.594	0.489	0.590	0.554	0.539	0.621	0.505	0.605	0.399	0.441	0.606
CH	0.647		0.583	0.642	0.594	0.308	0.203	0.244	0.263	0.445	0.510	0.552
FD1	0.583	0.618		0.262	0.462	0.511	0.560	0.493	0.529	0.367	0.358	0.422
FD2	0.638	0.652	0.383		0.473	0.490	0.614	0.500	0.588	0.463	0.443	0.501
HT	0.600	0.571	0.526	0.535		0.470	0.579	0.495	0.558	0.477	0.451	0.568
GrMD	0.610	0.371	0.575	0.533	0.522		0.270	0.049	0.364	0.452	0.459	0.551
RMD	0.667	0.198	0.604	0.629	0.562	0.348		0.245	0.212	0.518	0.548	0.478
YMD	0.588	0.319	0.560	0.558	0.534	0.157	0.333		0.342	0.461	0.467	0.548
Niño	0.651	0.270	0.584	0.606	0.555	0.404	0.241	0.403		0.469	0.527	0.517
PT	0.497	0.487	0.460	0.520	0.534	0.514	0.554	0.521	0.517		0.184	0.565
PTCR	0.532	0.524	0.466	0.496	0.526	0.522	0.568	0.523	0.541	0.313		0.553
RS	0.647	0.521	0.509	0.556	0.578	0.586	0.467	0.590	0.506	0.597	0.546	

Beach, FL, were all homogeneous (presumably due to their at least half-sib status) like the two homogeneous Costa Rican 'Atlantic Talls' AT02 and AT09 (all brown, Fig. 1a). The use of 'Atlantic Tall' samples from different origins helped to identify a distinct genetic group for this cultivar.

Regarding the six 'Panama Tall' genotypes used by Meerow et al. (2003) (PT01–PT06, lanes two to

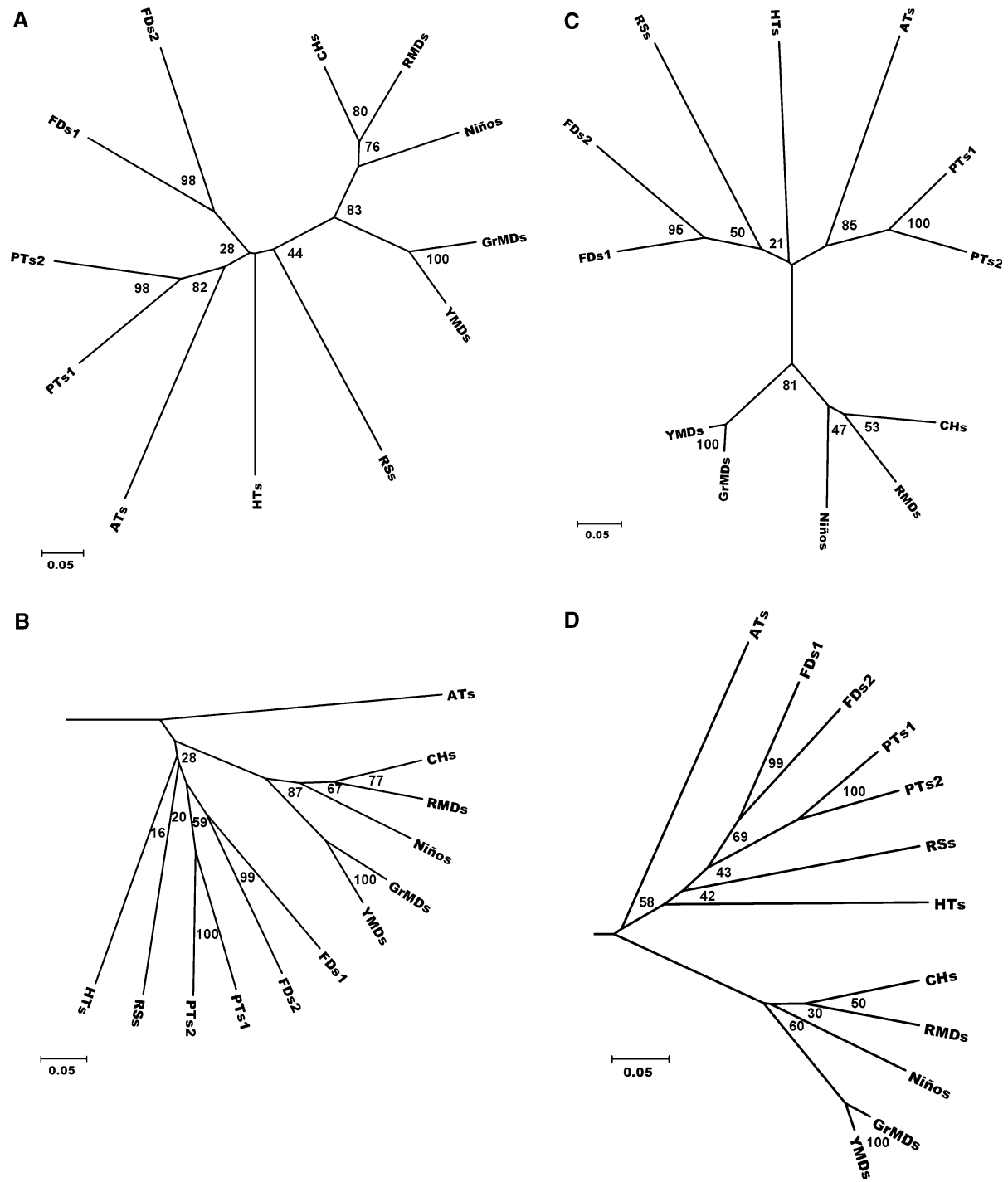

Fig. 2 Dendrograms depicting genetic relationships between ten coconut cultivars. **a**, **b** Cavalli-Sforza and Edwards' (1967) chord distance (Dc). **c**, **d** Nei et al.'s (1983) Da genetic distance. **a**, **c** Neighbor-joining (NJ, Saitou and Nei 1987). **b**, **d** Unweighted pair group method with arithmetic mean (UPGMA, Sneath and Sokal 1973). Numbers at tree nodes are bootstrap support percentages

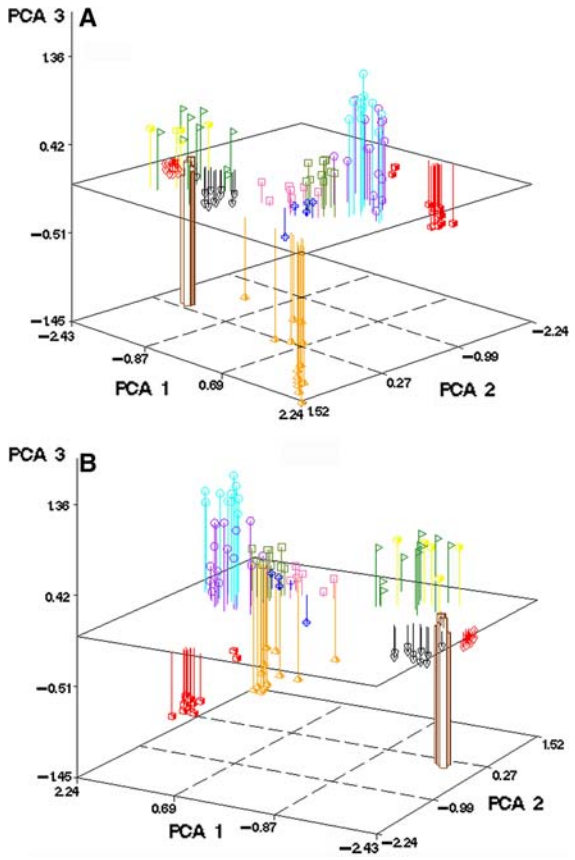

Fig. 3 Three dimensional PCA plots using Nei et al. (1983) Da genetic distance across 132 individual coconut genotypes representing ten varieties. **a**, **b** Depict the same plot from different viewing angles. 'Atlantic Tall' orange pyramid, 'Chowghat' brown pillar, 'Fiji Dwarf 1' purple balloon, 'Fiji Dwarf 2' cyan balloon, 'Hawaiian Tall' blue cross, 'Green Malayan Dwarf' green flag, 'Red Malayan Dwarf' red diamond, 'Yellow Malayan Dwarf' yellow prism, 'Green Niño' black heart, 'Panama Tall 1' pink square, 'Panama Tall 2' olive-green square, 'Red Spicata' red prism. The percentage of the variation explained by the axes was 52.33, 18.99 and 9.64%, respectively, for a total of 80.97%

seven in the PTs section in Fig. 1a), all but PT01 showed considerable levels of admixture. Two individuals were admixed with 'Green' or 'Yellow Malayan Dwarf' cultivars, two with 'Atlantic Talls' while PT06 appeared as a mixture of cultivars. These seeds were obtained from the Coconut Industry Board (CIB), Jamaica. Concerning the ten 'Panama Tall' genotypes from a local source in South Florida (PT1a–PT10a, Fig. 1a), seven of them showed variable levels of admixture primarily with the 'Atlantic Tall' cultivar two were homogeneous like PT1 and JT03 (all gray) while PT8a emerged as a combination between the two 'Panama Tall' sub-types (gray and blue colors). In addition, the 'Panama Talls' from the pure grove on the Pacific coast of Costa Rica (PTCR1–PTCR11, Fig. 1a) appeared as a distinct homogeneous group (all blue).

Of our sampling of 13 'Hawaiian Tall' genotypes, two of them (HT01 and HT02, maintained in Miami) appeared as mislabeled 'Panama Tall' genotypes; two (HT12 and HT13) appeared as hybrids between 'Atlantic' and 'Panama Tall' cultivars; and four others were admixed with 'Panama Tall' or 'Malayan Dwarf' cultivars. Nevertheless a distinct genetic group was identified for this cultivar (purple in Fig. 1a).

The 'Green Niños' were not different from the 'Chowghat' and 'Red Malayan Dwarf' cultivars based on this initial STRUCTURE analysis (yellow in Fig. 1a). In addition, several of them appeared to be admixed with 'Atlantic Tall', which may explain the weak grouping of this cultivar with the 'Atlantic Tall'—'Panama Tall' cluster in Meerow et al. (2003). A closer examination of the information received from the CIB, Jamaica, indicated that some of them could have been hybrids. The removal of the admixed individuals in the second STRUCTURE analysis led to the identification of a unique genetic group for this cultivar (yellow, Fig. 1b) and to its definite placing with the 'Chowghat' and the 'Malayan Dwarf' cultivars in the clustering analyses (Figs. 2, 3).

The results of the initial STRUCTURE analysis were used to remove the 86 hybrid/admixed genotypes and to reassign those individuals that evidently belonged to a different cultivar. The second STRUCTURE analysis with 132 individuals representing all ten coconut cultivars (Fig. 1b) also identified nine genetic groups, but with some differences from the results of the first analysis. As mentioned previously, removal of the putative hybrid 'Green Niños' resulted in the identification of a unique genetic group for this cultivar, while the number of genetic groups in the 'Panama Tall' cultivar was reduced from two to one, though we kept the two 'Panama Tall' subgroups for further analysis. The final list for cultivars and subgroups further studied based on the STRUCTURE analyses is presented in Table 3.

On a per locus basis, allele number ranged from two (CAC11, CAC13, CAC21, CAC27, CnW02, CnW06, CnW09 and CnW19) to 14 (CN2A5 and CNZ26) while average gene diversity ranged from

0.035 (CnW19 with two alleles) to 0.546 (CNZ02 with ten alleles). Polymorphism information content (PIC) values ranged from 0.091 (CnW02) to 0.838 (CNZ29) with a mean of 0.531 (Table 2). On a per cultivar basis the average allelic richness ranged from 1.01 ('Chowghat') to 1.90 ('Atlantic Tall') (Table 3). The two cultivars with the smaller average allelic richness and proportion of polymorphic loci values were 'Chowghat' and 'Red Malayan Dwarf' (Table 3). The two cultivars with the larger average allelic richness and proportion of polymorphic loci values were 'Atlantic Tall' and 'Fiji Dwarf'-1 (Table 3). Accordingly, these same cultivars ('Chowghat' plus 'Red Malayan Dwarf' and 'Atlantic Tall' with 'Fiji Dwarf'-1) had the lowest and highest gene diversity estimates, respectively. The 'Green Malayan Dwarf' cultivar had the highest fixation index (0.648), followed by the 'Panama Tall' and 'Green Niño' cultivars. 'Fiji Dwarf'-2, 'Hawaiian Tall', 'Panama Tall'-Costa Rica and 'Red Spicata' had higher observed heterozygosity values than expected, which led to negative fixation indexes (Table 3).

The largest genetic distance estimates were those between the 'Atlantic Tall' cultivar and each one of the other cultivars; the lower of these estimates were with the two 'Panama Tall' groups (Table 4). High genetic distance estimates were also obtained between the 'Fiji Dwarf' and the other cultivars, and between 'Red Spicata' and several of the other cultivars (Table 4). The dwarf cultivars overall ('Chowghat', 'Green Niño', 'Green', 'Red' and 'Yellow Malayan Dwarf') had the lowest genetic distance estimates among them (Table 4).

The clustering of the 'Chowghat', 'Red Malayan Dwarf', 'Green Niño', 'Green' and 'Yellow Malayan Dwarf' cultivars was consistent across clustering methods and genetic distance estimates with significant bootstrap values (Fig. 2). The clustering of the tall cultivars varied depending on the algorithm used. The NJ method (regardless of the genetic distance estimate) placed together the 'Atlantic Tall' cultivar and the two 'Panama Tall' genetic groups with significant bootstrap values (82 and 85%). However, the UPGMA method places the two 'Panama Tall' groups with the two 'Fiji Dwarf' groups with bootstrap values of 59 and 69%, and with the Da estimate it places the 'Atlantic Tall' cultivar in the 'Panama Tall'-'Fiji Dwarf' cluster with a bootstrap value of 58%. UPGMA with the Dc (Cavalli-Sforza and Edwards) estimate does not cluster the 'Atlantic Tall' cultivar with any of the other cultivars (bootstrap value of 28%). The NJ method with the Dc estimate also clustered the two 'Fiji Dwarf' groups with the tall cultivars but the bootstrap value was not significant (28%). The clustering of the 'Hawaiian Tall' and 'Red Spicata' cultivars was not significant with any of the genetic distance estimates and clustering methods (bootstrap support values $\leq$50%) though with the UPGMA method they were placed in the 'Fiji Dwarf'-'Panama Tall' cluster (Fig. 2). In addition to the higher bootstrap values for the cultivar clusters resolved by NJ, the R^2 values for the NJ trees were higher than those for UPGMA for both distance coefficients (Da: 88.7% vs. 79.9%; Dc: 92% vs. 88.4%).

The clustering of the 'Chowghat', 'Red Malayan Dwarf', 'Green Niño', 'Green' and 'Yellow Malayan Dwarf' cultivars was also observed in the PCA analysis (Fig. 3a), and except for the 'Green' and 'Yellow Malayan Dwarf' cultivars, each one of them formed a distinct cluster (Fig. 3b). The PCA analysis also supports the clustering of the two 'Panama Tall' genetic groups with the two 'Fiji Dwarf' genetic groups as the UPGMA clustering method did. In addition, the 'Hawaiian Tall' individuals were placed very close to the two 'Panama Tall' genetic groups by the PCA analysis. Finally, the 'Atlantic Tall' and 'Red Spicata' cultivars formed their own independent clusters, though the cluster closest to the 'Atlantic Talls' is that of the 'Hawaiian Tall'-'Panama Tall' group (Fig. 3a, b).

Discussion

Seed source, sample size and admixture levels

The initial analysis with STRUCTURE was quite useful in pin pointing those individuals which were highly admixed. This information was particularly helpful for the 'Atlantic Tall' and 'Panama Tall' cultivars since hybridization between them and also with other cultivars is a probable issue in germplasm collections. Our results confirm cross pollination between dwarf and tall types, which has been reported as common (Whitehead 1965; Harries 1978). Considerable levels of admixture as well as mislabeling were quite common in some of our

samples for our tall cultivars. The presence of 'Yellow Malayan Dwarf' or 'Atlantic Tall' alleles in 'Panama Tall' individuals from Jamaica was also detected by Baudouin et al. (2008). The level of admixture of the 'Atlantic Tall' and 'Panama Tall' genotypes used by Meerow et al. (2003), between them and with the same other cultivars, and the presence of one true to type 'Panama Tall' in each group, explains the clustering of these two cultivars in the neighbor joining tree and PCA plot corresponding to Figs. 1 and 3 in Meerow et al. (2003).

The increased number of individuals sampled in this study, particularly for the 'Atlantic Tall' and 'Panama Tall' cultivars, helped in identifying unique genetic groups for these cultivars as well as in the establishment of more accurate relationships between them by eliminating admixed individuals. Many coconut studies have not accounted for this factor in their analyses and have evaluated just between one and three (maximum seven) individuals for most of their cultivars (Manimekalai and Nagarajan 2006a, b; Perera et al. 1998, 1999, 2000; Rivera et al. 1999; Teulat et al. 2000) which may be a critical factor when evaluating relationships between tall (outcrossing) genotypes. Furthermore, only recently, Baudouin and Lebrun (2009), Baudouin et al. (2008) and Lebrun et al. (2008) assessed more than 20 and up to 104 individuals for specific cultivars ('Malayan Yellow Dwarf', 'Panama Tall') and showed the prevalence of cross-contamination/admixture in those cutlivars.

The analysis of a number of 'Panama Tall' individuals by Lebrun et al. (2005) and Baudouin et al. (2008) led to the identification of three sub-populations: the "typical" 'Panama Tall', the "Aguadulce" and the "Costa Rica" types. Those studies identified the "Aguadulce" type as having a low percentage of 'Atlantic Tall' genes. Our results indicate that we might have had some of the "Aguadulce" type individuals (Fig. 1a). However, to truthfully determine the relationship between these two cultivars we removed these individuals from our final analysis. Our final sampling probably included only the "typical" 'Panama Tall' and the "Costa Rica" types.

In general, higher levels of admixture were observed with the tall cultivars as compared to the dwarf cultivars, as expected due to their out-crossing nature. On the other hand, the majority of the dwarf cultivars showed fewer admixed individuals, regardless of their source, particularly the 'Chowghat' and the 'Malayan Dwarf' cultivars. In addition, some of the 'Red Spicata' individuals showed some level of admixture with the 'Green Niño', 'Malayan Dwarf' or 'Panama Tall' cultivars, but a unique genetic group was also obvious for this cultivar (red in Fig. 1a).

The 'Green Niños' presented an interesting situation as several appeared to be introgressed with 'Atlantic Tall' (Fig. 1a). When we included putative hybrid individuals in the population sample, the placement of this cultivar in our initial preliminary dendrograms was with the 'Atlantic Talls' (data not shown), and is indicative of how profoundly undetected admixture in population samples can influence the topology of genetic distance trees.

The exception to high homogeneity in dwarf coconuts is the 'Fiji Dwarf' or 'Niu Leka' cultivar. Both Harries (1978) and Whitehead (1976) consider this variety more like a tall variety in most respects other than the fact that it segregates for phenotypes with compressed internodes. The reproductive biology of this cultivar (conclusion of the male phase one to five days prior to the beginning of the female phase within an inflorescence) favors cross-pollination (Whitehead 1966). The representatives of this cultivar were partitioned in two groups, and their overall level of admixture denotes their outcrossing behavior, although the homogeneity of several of them is striking (all green or pink color, Fig. 1). Five of the 'Fiji Dwarf' individuals imported from Jamaica in the 1980 s (UF14–UF17 and FD true), which have never succumbed to Lethal Yellowing in South Florida, were highly homogeneous. Several progeny from these five original 'Fiji Dwarf' individuals were also highly homogeneous, while others showed admixture with other cultivars ('Atlantic Tall', 'Malayan Dwarf', 'Panama Tall') maintained in close proximity.

Method of analysis and coconut cultivar relationships

NJ clustering differs from UPGMA by not assuming equal rates of evolution across all lineages, and is considered a more accurate clustering method for genetic distances (Saitou and Nei 1987). Except for studies by Perera et al. (1998, 2000, 2003), most of the other coconut studies on genetic diversity have used the UPGMA method (e.g., Ashburner et al. 1997; Rivera et al. 1999; Teulat et al. 2000;

Manimekalai and Nagarajan 2006a, b) to present the relationships between cultivars [Lebrun et al. (1998) used Factorial Analysis of Correspondences (FAC) instead of cluster analysis]. In addition, some of these studies have used as few as one to three individuals per cultivar in their analysis, which according to our results and those of Baudouin et al. (2008), may cause misleading resolution of relationships between cultivars due to the presence of admixed individuals.

The grouping of the cultivars with the two clustering methods used (NJ and UPGMA) showed some consistent groupings but also some remarkable differences (Fig. 2). The cluster including the five dwarf cultivars: 'Chowghat', 'Green Niño', 'Green', 'Red' and 'Yellow Malayan Dwarf' was quite consistent with both clustering methods and genetic distance estimates (Fig. 2) and was in agreement with their low genetic distance estimates (Da and Dc, Table 4). The support (bootstrap values) for almost every node of this main cluster was equal or higher than 50%, the exception being the node for the 'Green Niño' and the 'Chowghat'—'Red Malayan Dwarf' cultivars (bootstrap value >60% with Dc but <50% with Da). This same cluster is also observed in the PCA plots (Fig. 3) where, with the exception of the 'Green' and 'Yellow Malayan Dwarf' cultivars, a discrete group can be observed for each one of the other three cultivars (Fig. 3b). The grouping of dwarf cultivars has been consistently observed in other studies and has been attributed to a common origin (Lebrun et al. 1998; Perera et al. 1998, 2003; Rivera et al. 1999; Teulat et al. 2000; Upadhyay et al. 2004). The clustering of the 'Green' and 'Yellow Malayan Dwarf' cultivars is easily explained by their shared origin while the closer relationship of the 'Red Malayan Dwarf' with the 'Chowghat' and the 'Green Niño' cultivars rather than with the other 'Malayan Dwarf' cultivars was unexpected. Harries (1978) placed the 'Chowghat' dwarf cultivar (native to India) in the same group of small-fruited dwarf types as the 'Green Niño' (i.e. 'Coconino', from the Philippines). However, he identified the 'Malayan Dwarf' cultivars in a different group due to their larger-fruit and as a "more vigorous" dwarf type (Harries 1978). Our results contradict such grouping but also corroborate the closeness between the 'Chowghat' and the 'Green Niño' cultivars. The pairing/clustering of 'Chowghat Dwarf' and 'Malayan Dwarf' cultivars was also identified by Upadhyay et al. (2004) and by Devakumar et al. (2006). This study corroborates a common ancestry/domestication region for dwarfs from India, Malaysia and the Philippines as established/proposed in previous studies (Harries et al. 2004; Lebrun et al. 1998, 2005; Perera et al. 2003; Teulat et al. 2000).

The pairing/clustering of the 'Atlantic Tall' and the two 'Panama Tall' cultivars was robust and consistent with the NJ method (bootstrap values > 80%) but inconsistent with the UPGMA method (Fig. 2). UPGMA instead placed the two 'Panama Tall' genetic groups with the two 'Fiji Dwarf' genetic groups (bootstrap values of 59 and 69%) although with the Da estimate 'Atlantic Tall' was grouped with the 'Panama Tall'-'Fiji Dwarf' cluster (bootstrap value of 58%). 'Atlantic Tall' was not part of any cluster by using the Dc estimate and UPGMA. The groups closest to the 'Atlantic Talls' in the PCA plot are the 'Hawaiian Tall', the two 'Panama Tall' sub-groups and the 'Fiji Dwarfs' (Fig. 3). This clustering is consistent with the low genetic distance values between 'Atlantic Tall' and the aforementioned cultivars (Table 4).

The coconut in the Americas has been long proposed as having two sources: one from across the Pacific ocean that led to the Gulf of Panama as a main center of distribution, which existed prior to the arrival of Europeans, and the second from the Cape Verde islands from where coconuts were brought to Puerto Rico (Bruman 1944). Bruman (1944) also dismissed the only report by Columbus of tall palms with large nuts on the north coast of Cuba as a possible fabrication by the editor of his notes. Harries (1977) further explained the source of the Cape Verde coconuts as being introduced from India and East Africa, believed to be the center of domestication of those cultivars growing on the Atlantic and the Caribbean, including the 'Atlantic Tall' from our study. On the other hand, the origin of the 'Panama Tall' cultivar (from the Pacific American coast) has recently been confirmed as South–East Asia (another center of domestication of coconut), particularly the Philippines (Baudouin and Lebrun 2009). The proposed origins of these two cultivars, and their contrasting phenotypic characteristics, make the explanation of their well-supported pairing by the NJ method in our study confounding. The theory about the Indian and East African origin for the cultivars from the Atlantic and the Caribbean in America has been widely championed (Harries 1978; Lebrun et al. 2005; Perera et al. 2000, 2003;

Purseglove 1985), but little support has been provided with genetic marker data. Only recently, a genetic study of the 'Atlantic Tall' ('Jamaica Tall') and 'Panama Tall' cultivars along with African and Caribbean coconuts reported dissimilar results depending on the marker system used (Manimekalai and Nagarajan 2006a, b). The RAPD analysis (Manimekalai and Nagarajan 2006a) paralleled the accepted origin theory by grouping the 'Atlantic Tall' with an African (Nigerian) and Indian cultivars in one main cluster and the 'Panama Tall' with cultivars from the South Pacific and South East Asia in another (though 'Panama Tall' was paired with a cultivar from North East India). The Nigerian cultivar was genetically most similar to a Caribbean cultivar. Alternatively, the ISSR study (Manimekalai and Nagarajan 2006b) found highest similarity between 'Panama Tall' and the Nigerian cultivar, placing them both in a cluster that included the 'Atlantic Tall' and Caribbean cultivars as well as other Indian and South East Asian genotypes. Both studies used the UPGMA clustering method and only the ISSR results were reported with bootstrap values. Such discrepancies with the same genotypes make it difficult to conclude that any one scenario of coconut varietal relationships is accurate. Zizumbo-Villarreal et al. (2006) studied Atlantic and Pacific genotypes from Mexico and analyzed them with both clustering methods (UPGMA and NJ). Their results were similar to ours; with NJ tall and dwarf cultivars were separated into two main groups, though their 'Malayan Dwarfs' were grouped with some of their 'Panama Talls' and, with some exceptions, their talls were divided between those of Atlantic origin and those from the Pacific. With UPGMA their 'Malayan Dwarfs' also grouped with a subgroup of 'Pacific Talls' and the Atlantic and Pacific talls are grouped in two separate clusters. The fact that they found a consistent grouping of their 'Malayan Dwarfs' with a subgroup of 'Pacific Talls' may be indicative of admixture between those Pacific genotypes, which, as we see with our own data, can play a significant role in the clustering patterns of the different genetic groups in the dendrograms.

In summary, broad-based molecular marker studies of coconut varietal relationships should strive to include at least 10–20 individuals from at least two isolated source populations if at all possible. More importantly, it is essential that population samples be tested for possible admixture with other varieties growing in their vicinity.

Whether recent cross contamination or an earlier dispersal of Pacific tall coconuts into the Atlantic area led to a much closer relationship than expected between our 'Atlantic Tall' and 'Panama Tall' materials remains to be further investigated. Given the fact that NJ clustering results portray different and even better supported grouping results than UPGMA might be indicative that the ancestry of coconut cultivars may be more ambiguous than presumed by some accounts (e.g., Harries 1978; Lebrun et al. 2005). At the very least, molecular marker studies on coconut diversity should provide bootstrap support percentages for any genetic distance tree used to justify a particular classification, as well as a rationale for rejecting alternative hypotheses of relationships based on other clustering algorithms. Common ancestry of dwarf cultivars on one hand and of tall cultivars on another as resolved by NJ clustering with relatively high bootstrap support may suggest that a different model for the domestication and dispersal of coconut cultivars may need to be examined. A phylogeographic approach (Avise 2000, 2009) using multilocus sequence data across a large and geographically diverse collection of coconut populations may ultimately be more informative in answering this question decisively than any method of fragment analysis.

References

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

Avise JC (2000) Phylogeography: the history and formation of species. Harvard University Press, Cambridge, MA

Avise JC (2009) Phylogeography: retrospect and prospect. Journal of Biogeography 36:3–15

Baudouin L, Lebrun P (2009) Coconut (*Cocos nucifera* L.) DNA studies support the hypothesis of an ancient Austronesian migration from Southeast Asia to America. Genet Resour Crop Evol 56:257–262

Baudouin L, Lebrun P, Berger A, Myrie W, Been B, Dollet M (2008) The Panama Tall and the Maypan hybrid coconut in Jamaica: did genetic contamination cause a loss of resistance to lethal yellowing? Euphytica 161:353–360

Botstein D, White RL, Skolnick M, Davis RW (1980) Construction of a genetic linkage map in man using restriction fragment length polymorphisms. Am J Hum Genet 32:314–331

Bruman HJ (1944) Some observations on the early history of the coconut in the new world. Acta Americana 2:220–243

Cavalli-Sforza LL, Edwards AWF (1967) Phylogenetic analysis: models and estimation procedures. Amer J Hum Genet 19:233–257

Devakumar K, Jayadev K, Rajesh MK, Chandrasekar A, Manimekalai R, Kumaran PM, Parthasarathy VA (2006) Assessment of the genetic diversity of Indian coconut accessions and their relationship to other cultivars, using microsatellite markers. Plant Gen Res Newsletter FAO-Bioversity 145:38–45

El Mousadik A, Petit RJ (1996) High level of genetic differentiation for allelic richness among populations of the argan tree *Argania spinosa* (L.) Skeels endemic to Morocco. Theor Appl Genet 92:832–839

Evanno G, Regnaut S, Goudet J (2005) Detecting the number of clusters of individuals using the software structure: a simulation study. Molecular Ecology 14:2611–2620

Falush D, Stephens M, Pritchard JK (2003) Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies. Genetics 164:1567–1587

Glaubitz JC (2004) Convert: a user-friendly program to reformat diploid genotypic data for commonly used population genetic software packages. Mol Ecol Notes 4:309–310

Goudet J (2002) FSTAT (vers. 2.9.3.2), a program to estimate and test gene diversities and fixation indices. Available from http://www.unil.ch/izea/softwares/fstat.html

Harries HC (1977) The cape verde region (1499–1549); the key to coconut culture in the western hemisphere? Turrialba 27:227–231

Harries HC (1978) The evolution, dissemination and classification of *Cocos nucifera* L. Bot Rev 44:265–320

Harries HC (2001) The coconut palm (*Cocos nucifera* L.). In: Last FT (ed) Tree crop ecosystems. Elsevier, New York, pp 321–338

Harries H, Baudouin L, Cardeña R (2004) Floating, boating and introgression: Molecular techniques and the ancestry of coconut palm populations on Pacific islands. Ethnobotany Res Appl [online] 2:37-53 http://128.171.206.29/ojs/index.php/era/article/view/36/24

Kalinowski ST (2007) TreeFit, a computer program for evaluating how well evolutionary trees fit genetic distance data. Available for download at http://www.montana.edu/kalinowski/Software/TreeFit2.htm

Langella O (2002) Populations 1.2.28: population genetic software (individuals or populations distances, phylogenetic trees). Available from http://www.cnrs-gif.fr/pge

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

Lebrun P, Berger A, Hodgkin T, Baudouin L (2005) Biochemical and molecular methods for characterizing coconut diversity. In: Batugal P, Ramanatha R, Oliver J (eds) Coconut genetic resources. International plant genetic resources Institute—regional office for Asia, the Pacific and Oceania (IPGRI-APO). Serdang, Selangor, pp 225–247

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

Lewis PO, Zaykin D (2001) GENETIC DATA ANALYSIS: computer program for the analysis of allelic data. Version 1.1. Free program distributed by the authors over the internet from http://hydrodictyon.eeb.uconn.edu/people/plewis/software.php

Liu K, Muse SV (2005) Powermarker: Integrated analysis environment for genetic marker data. Bioinformatics 21:2128–2129

Manimekalai R, Nagarajan P (2006a) Interrelationships among coconut (*Cocos nucifera* L.) accessions using RAPD technique. Genet Resour Crop Evol 53:1137–1144

Manimekalai R, Nagarajan P (2006b) Assessing genetic relationships among coconut (*Cocos nucifera* L.) accessions using inter simple sequence repeat markers. Sci Hort 108:49–54

Mauro-Herrera M, Meerow AW, Borrone JW, Kuhn DN, Schnell RJ (2006) Ten informative markers developed from WRKY sequences in coconut (*Cocos nucifera*). Mol Ecol Notes 6:904–906

Mauro-Herrera M, Meerow AW, Borrone JW, Kuhn DN, Schnell RJ (2007) Usefulness of WRKY gene-derived markers for assessing genetic population structure: an example with Florida coconut cultivars. Sci Hort 115:19–26

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 Appl Genet 106:715–726

Nei M (1987) Molecular evolutionary genetics. Columbia University Press, New York

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 Appl Genet 96:545–550

Perera L, Russell JR, Provan J, Powell W (1999) Identification and characterization of microsatellite loci in coconut (*Cocos nucifera* L.) and the analysis of coconut populations in Sri Lanka. Mol Ecol Notes 8:344–346

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

Petit RJ, El Mousadik A, Pons O (1998) Identifying populations for conservation on the basis of genetic markers. Conserv Biol 12:844–855

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

Purseglove JW (1985) Tropical crops: monocotyledons, 5th edn. Longman, London

Rivera R, Edwards KJ, Barker JHA, Arnold GM, Ayad G, Hodgkin T, Karp A (1999) Isolation and characterization of polymorphic microsatellites in *Cocos nucifera* L. Genome 42:668–675

Rohde W, Kullaya A, Rodriguez J, Ritter E (1995) Genome analysis of *Cocos nucifera* L. by PCR amplification of spacer sequences separating a subset of copia-like *Eco*R1 repetitive elements. J Genet Breed 49:179–186

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

Saitou N, Nei M (1987) The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol Biol Evol 4:406–425

Sneath PHA, Sokal RR (1973) Numerical taxonomy: the principles and practice of numerical classification. WH Freeman, San Francisco

Takezaki N, Nei M (1996) Genetic distances and reconstruction of phylogenetic trees from microsatellite DNA. Genetics 144:389–399

Teulat B, Aldam C, Trehin R, Lebrun P, Barker JHA, Arnold GM, Karp A, Baudouin L, Rognon F (2000) An analysis of genetic diversity in coconut (*Cocos nucifera*) populations from across the geographic range using sequence-tagged microsatellites (SSRs) and AFLPs. Theor Appl Genet 100:764–771

Upadhyay A, Jayadev K, Manimekalai R, Parthasarathy VA (2004) Genetic relationship and diversity in Indian coconut accessions based on RAPD markers. Sci Hort 99:353–362

Van den Berg RG, Jacobs MMJ (2007) Molecular taxonomy. In: Vreugdenhil D (ed) Potato biology and biotechnology. Elsevier, New York, pp 55–76

Whitehead RA (1965) Speed of germination, a characteristic of possible taxonomic significance in *Cocos nucifera* L. Trop Agric (Trinidad) 42:369–372

Whitehead RA (1966) Sample survey and collection of coconut germplasm in the Pacific Islands (30 May–5 September 1964). Ministry of Overseas Development, HMSO, London

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

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