

# Genetic Characterization of Ghanaian Avocados Using Microsatellite Markers

Atiako Kwame Acheampong<sup>1,2</sup>, Richard Akromah, and Felix A. Ofori

Department of Crop and Soil Sciences, Kwame Nkrumah University of Science and Technology (KNUST), Kumasi, Ghana

Jemmy F. Takrama

Physiology/Biochemistry Division, Cocoa Research Institute of Ghana (CRIG), P.O. Box 8, New Tafo-Akim, Ghana

David Saada, Iris Bitton, and Uri Lavi

Department of Fruit Trees Sciences, Institute of Plant Sciences, ARO–Volcani Center, P.O. Box 6, Bet-Dagan 50250, Israel

ADDITIONAL INDEX WORDS. founder parents, diversity, horticultural races, *Persea americana*, phylogenetic tree, SSR

**ABSTRACT.** The origin, horticultural race and genetic relationship among Ghanaian avocado (*Persea americana* Mill.) seedlings were studied. A population of 172 avocado samples, selected from farmers' fields and backyards, were genotyped using 12 simple sequence repeat markers. Representatives of the three horticultural races served as outliers in this study. An average 4.42 alleles per locus were elucidated across all populations. The average expected heterozygosity,  $H_e$  (0.51), was 41.7% greater than the observed heterozygosity,  $H_o$  (0.36), depicting heterozygote deficiency and a high level of homogeneity or inbreeding within the entire Ghanaian population. The genetic distances between the Ghanaian subpopulation pairs was low (0.00 to 0.07), further supporting a high level of homogeneity and suggesting either very few founder parents or a high proportion of founders with very little genetic variability. The phylogenetic tree generated reaffirmed the high genetic similarity among the Ghanaian seedlings with only 8.7% clustering separately. The Ghanaian population also was found to be genetically more closely related to the West Indian representatives but distinct from the Mexican and the Guatemalan accessions, proposing that either most of the founders may have been of the West Indian race or founders of the other races may not have survived the harsh tropical conditions of Ghana. It is suggested that enlarging Ghana's avocado gene pool by introducing various elite cultivars will improve the avocado industry in this country.

Avocado is the only edible member of the small yellowish flower-bearing plant family, the Lauraceae (Morton, 1987), which includes mostly tropical or subtropical plants. The fruit is believed to have originated from the eastern and central highlands of Mexico, to the Pacific coast of Central America, from where it was distributed to other parts of the world (Popenoe, 1920; Storey et al., 1986). Several reviews detailing various aspects of the avocado history and distribution have been published (Knight, 2002; Popenoe and Zentmyer, 1997).

The three horticultural races of avocado, designated as Mexican [*P. americana* var. *drymifolia* (Schecht. and Cham.) Blake], West Indian (*P. americana* var. *americana* Mill.), and Guatemalan (*P. americana* var. *guatemalensis* Williams) (Scora and Bergh, 1990), believed to be evolutionary products of the primitive *P. americana*, became botanical types when they developed in geographical isolation from one another (Popenoe, 1920; Storey et al., 1986). The unique protogynous and diurnally synchronous dichogamous mechanism exhibited

by avocado flowers (Stout, 1923) coupled with the absence of sterility barriers among the horticultural races has resulted in great diversity and polymorphism in avocado.

The fruit, now cultivated almost exclusively in the forest regions of Ghana as interplants on farmers' fields (A.K. Acheampong, R. Akromah, F.A. Ofori, and J.F. Takrama, unpublished data), was first introduced into the country during the precolonial times by the missionaries with peasant production not beginning until 1907 (Ghana Division of Agriculture, 1961). Improved commercial cultivars, mostly West Indian and West Indian × Guatemalan hybrids (including the cultivars Booth 7, Booth 8, Fuchsia, Lula, Monroe, Trapp, Choquette, Collinson, and Waldin), were introduced into the country later by the Plant Protection and Quarantine unit of U.S. Department of Agriculture–Animal and Plant Health Inspection Services, and U.S. Agency for International Development (USAID). In the 1960s, 'Duke', 'Ettinger', and 'Fuerte', from the National Germplasm Repository (NGR) in Miami, FL, were among the latest cultivars to be introduced into the country. Today, with no certified nurseries to provide farmers with vegetative clones, all the country's avocados are solely propagated from seeds, which probably resulted from these founder parents.

Avocado genetic diversity studies have been conducted to assess variations among and within races and to classify interracial hybrids (Rhodes et al., 1971). These studies have also been aimed at establishing a genetic basis for the morphological classification of the individuals into these botanical/horticultural

Received for publication 20 Mar. 2008. Accepted for publication 11 Sept. 2008. This work was supported by a fellowship from the Kirkhouse Trust, UK.

We are grateful to the Directors and Management of Agricultural Research Organization (ARO), Israel, and the Cocoa Research Institute of Ghana (CRIG), Tafo, Ghana. We also thank Avraham Ben-Ya'acov, Emi Lahav, and Nana Pokua Frimpong for their useful suggestions.

<sup>1</sup>Current address: Department of Fruit Trees Sciences, Institute of Plant Sciences, Agricultural Research Organization–Volcani Center, P.O. Box 6, Bet-Dagan 50250, Israel.

<sup>2</sup>Corresponding author. E-mail: ak\_acheampong@yahoo.com.

racess and for paternity identification. DNA markers such as random amplified polymorphic DNA (Chang et al., 2003; Fiedler et al., 1998), restriction fragment length polymorphism (Davis et al., 1998), and minisatellites (Lavi et al., 1991; Mhameed et al., 1995) have been widely applied to such studies. Being more abundant, polymorphic, codominant, and evenly distributed in the genome, microsatellites have gained immense popularity in genetic mapping, genotyping, and genetically grouping avocado accessions into the various races (Lavi et al., 1994; Mhameed et al., 1996, 1997; Schnell et al., 2003; Sharon et al., 1997). These strides notwithstanding, very little is known about the genetic diversity within the samples grown in Ghana. Biodiversity study using amplified fragment length polymorphism, conducted by Taah et al. (2003), was incomplete and inconclusive.

The aim of this study was to assess the genetic diversity within and between the avocado subpopulations found in the six forest regions of Ghana and to further determine the racial classification of the selected Ghanaian samples by estimating their genetic relatedness to representatives of the West Indian, Guatemalan, and Mexican races.

### Materials and Methods

**PLANT MATERIALS.** One hundred seventy-two Ghanaian avocado samples and 24 avocado cultivars from the three horticultural races (Table 1) were analyzed in this study. The Ghanaian seedlings were collected mostly from farmers' fields and backyards in the six forest regions of the country (Fig. 1), whereas the racial representatives were collected from the germplasm repository of the Agricultural Research Organization, Bet Dagan, Israel. In accordance with reports by Ben-Ya'acov (1995), but contrary to that of Schnell et al. (2003), we considered 'Orizaba 3' as West Indian.

**DNA EXTRACTION.** Genomic DNA was isolated from fresh, moderately young leaves of the previously described plant materials by a modified cetyltrimethylammonium bromide method (Russell et al., 1992). The DNA was quantified by agarose gel electrophoresis using the  $\lambda$  DNA-*Hind III* digest ladder (Invitrogen, Carlsbad, CA) as the reference.

**SIMPLE SEQUENCE REPEAT MARKERS.** The simple sequence repeat (SSR) primers used in this study (Table 2) were

Table 1. The names and racial lineages of the 24 known avocado cultivars analyzed by microsatellite markers in this study.

Racial background of accessions <sup>a</sup>		
Mexican	West Indian	Guatemalan
Aguacatitlan	Antigua	Amatenango
Aquila 1	Bernercker	Apakia 2
Aquila 2	Butler	CH-C11 Padua
Banos	DW1 Bank	Comitan 1
Basaldua	Fuchsia	Las Americas
	Galvan	Nabal
	Orizaba 1	Puerto Viejo
	Orizaba 3	Reed
	Tantima	
	Tower 2	
	Waldin	

<sup>a</sup>Accessions were selected on the expert advice of A. Ben-Ya'acov and E. Lahav, both formerly of the Department of Fruit Trees, Agricultural Research Organization, Israel.

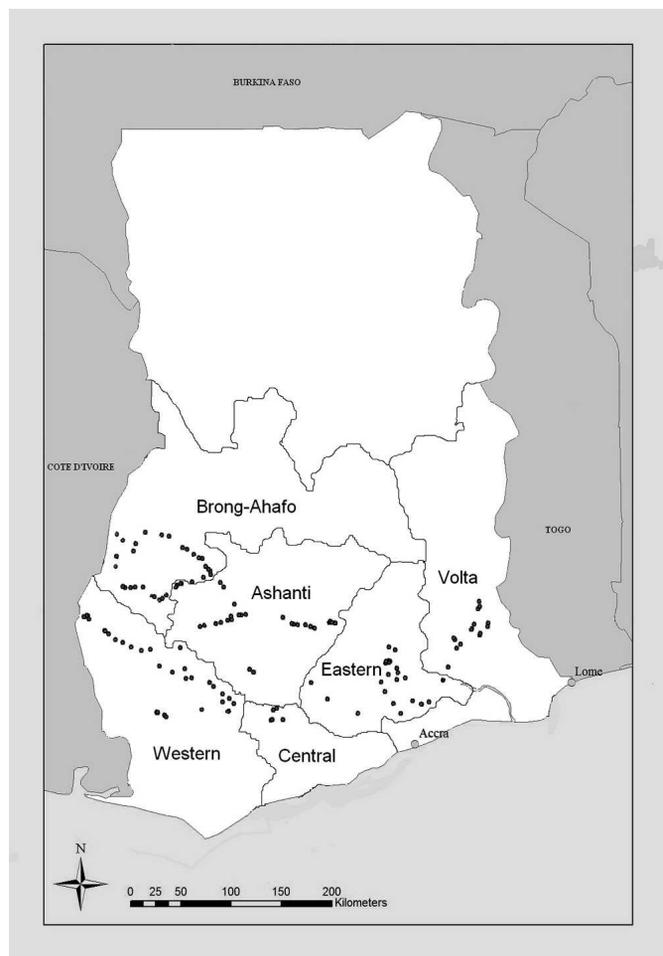


Fig. 1. Map of Ghana indicating the sampling sites of the avocado seedlings analyzed in this study. The samples were collected from the six forest regions of the country: Ashanti, Brong-Ahafo, Central, Eastern, Volta, and Western regions; ● = sample collection point.

generated and used by Sharon et al. (1997) and later applied by Schnell et al. (2003).

**POLYMERASE CHAIN REACTION AMPLIFICATIONS.** Polymerase chain reaction (PCR) amplifications were performed in a total volume of 10  $\mu$ L containing 10 ng DNA, 0.5  $\mu$ M each of the forward and reverse primers, 0.2 mM of each of the four deoxyribonucleoside triphosphates, 1 $\times$  *Taq* buffer (containing 1.5 mM MgCl<sub>2</sub>, 50 mM Tris-HCl pH = 9, 50 mM KCl, 0.1% Triton X-100), 0.5 U *Taq* DNA polymerase (New England BioLabs, Ipswich, MA), and 0.1  $\mu$ L of 3000 Ci/mmol  $\alpha$ <sup>32</sup>P-dCTP (GE Healthcare, Little Chalfont, U.K.). PCR reactions were carried out on a DNA Engine thermal cycler (MJ Research, Watertown, MA) and consisted of the following profile: 3 min denaturation at 94 °C followed by 35 cycles of denaturation at 94 °C for 30 s, 60 s at the appropriate primer annealing temperature, and 60 s extension at 72 °C with a final 10 min extension at 72 °C.

**ELECTROPHORESIS.** The PCR product (10  $\mu$ L) was mixed with 10  $\mu$ L of loading solution (containing 95% formamide, 0.02 M EDTA pH = 8, 1% bromophenol blue, 1% xylene cyanol, 10 mM NaOH), heat-denatured, and electrophoresed on 6% denaturing polyacrylamide sequencing gel at 80 W, 2 kV for 2 h. Gels were dried at 80 °C for 2 h and exposed to KODAK Biomax MS X-ray films (Eastman Kodak, Rochester, NY) for

Table 2. Characteristics of the simple sequence repeat (SSR) loci and primers used for the molecular characterization of Ghanaian avocado seedlings.

SSR locus <sup>z</sup>	Repeat motif <sup>z</sup>	Primer sequence (5'-3') <sup>y</sup>	H <sub>o</sub> <sup>y</sup>	H <sub>e</sub> <sup>y</sup>	Homozygotes (no.) <sup>x</sup>	Heterozygotes (no.) <sup>x</sup>	Null alleles (no.) <sup>x</sup>	Departure from HWE <sup>w</sup>
AVAC01	(TG) <sub>15</sub>	F: CTGGTTGCTCTCTTGCTACATAATA R: CGGTTTTGTAAGTTGATAG	0.35	0.42	106	61	5	
AVAG03	(TC) <sub>17</sub>	F: GCACCTCCTAAACTTGCAGGT R: CTGAACATCCAAATGACAAACATCC	0.28	0.39	113	52	7	
AVMIX03	(TG) <sub>16</sub> , (AG) <sub>20</sub>	F: GATATTCCTGTTGCTACTGC R: AATGTTCCCCCATGAAAAGTCTCC	0.52	0.77	62	103	7	
AVMIX04	(AG) <sub>12</sub> , (CAA) <sub>5</sub> , (ACAG) <sub>10</sub>	F: CCGTTTGTCTTCCCTGTATC R: GTTATCCCTTCCACTTTC	0.49	0.65	74	90	8	
AVAG05	(AG) <sub>10</sub>	F: GGATCTGATGTGGGGGAG R: CCTGTCCGAAAAGACTATGCG	0.22	0.36	133	34	5	
AVAG06	(CT) <sub>18</sub>	F: CGACCTTCTTATACTC R: GTACCTCTGATAATGAGCAT	0.35	0.47	95	66	11	
AVAG07	(TC) <sub>15</sub>	F: ATCCAAAATGCACAAAGGTGAGG R: TGTCGCTATGTCCAAAATGTGG	0.48	0.50	86	78	8	**
AVAG10	(CT) <sub>22</sub>	F: GAATTACAAAGCAGCTAGAG R: GTAGAAAAGTGGGCACACAT	0.33	0.53	111	52	9	**
AVAG11	(AG) <sub>20</sub>	F: AGCGATGAACATTAACCCA R: ATTTCTTCAACCCCATCTGTC	0.21	0.39	123	42	7	
AVAG13	(CT) <sub>18</sub>	F: CTGCGATAACAACTGGAC R: AACTAGGACCTGAAACCCG	0.20	0.41	142	20	10	**
AVAG22	(GA) <sub>15</sub>	F: GATCATCAAGTCTCCTCTGG R: GATCTCATAGTCCAAAATAATGC	0.43	0.51	85	79	8	
AVAG25	(TC) <sub>14</sub>	F: ATGGTTTTTCTGCCCCCTT R: AACAAAGCCCCCTAAAAGAA	0.48	0.69	63	99	10	
Mean			0.36	0.51	99.42	64.67	7.91	
Total					1,193	776	95	

<sup>z</sup>Information obtained from Sharon et al. (1997).

<sup>y</sup>H<sub>o</sub> = observed heterozygosity; H<sub>e</sub> = unbiased expected heterozygosity (Levene, 1949).

<sup>x</sup>Parameters are based on the analysis of the Ghanaian avocado seedlings in this study.

<sup>w</sup>HWE = Hardy-Weinberg equilibrium; \*\*deviates significantly from HWE at *P* < 0.01.

4 to 10 h. Films were developed using Agfa Curix 60 X-ray developer (Agfa HealthCare, Mortsel, Belgium), and the sizes of the PCR products were estimated directly from the autoradiograph using a 10-basepair (bp)-sized standard (Invitrogen).

**DATA ANALYSIS.** Gene diversity for each locus and the average across all loci for all six Ghanaian subpopulations were calculated using Nei's (1978) unbiased estimate. Number of alleles and allelic frequencies for each SSR and across all populations were estimated using POPGENE version 1.31 (Yeh and Boyle, 1997). The expected ( $H_e$ ) and the observed ( $H_o$ ) heterozygosities were estimated from the allelic frequencies using GENEPOP version 3.4 (Raymond and Rousset, 2007). Phylogenetic trees, based on Cavali-Sforza and Edwards (1967) chord distances, were constructed by neighbor-joining (NJ) (Saitou and Nei, 1987) using MICROSAT 2, PHYLIP version 3.5 (Felsenstein, 1989) and TreeView version 1.6.6 (Page, 1996) to depict the genetic relatedness of the Ghanaian samples and accessions from the three horticultural races. For all phylogenetic trees constructed, bootstrap values were computed using 10,000 replications of the SSR genotyping data. The inbreeding coefficient ( $F_{IS}$ ) (Wright, 1978), which measures the heterozygosity deficiency or excess within a population, was estimated for each of the Ghanaian subpopulations. The  $\chi^2$  estimation of the Hardy-Weinberg (HW) hypothesis was performed by GENEPOP version 3.4 to assess the deviation of each locus from HW equilibrium. The genetic distances between subpopulation pairs were calculated following Nei (1978) using Tools for Population Genetic Analysis (TFPGA) version 1.3 (Miller, 1997).

## Results

**CHARACTERISTICS OF THE 12 SIMPLE SEQUENCE REPEATS.** The 12 SSRs applied in this research were polymorphic and generated 53 alleles across the 172 Ghanaian seedling population (Tables 2 and 3). The number of different alleles elucidated ranged from two (AVAG05) to seven (AVAC01, AVMIX03, and AVAG10), whereas sizes ranged from 64 bp (AVAG06) to 199 bp (AVAG10), and allelic frequencies varied between 0.3% and 86.8%. Across all 12 loci, samples were scored 1193 times (57.8%) as homozygotes and 776 times (37.6%) as heterozygotes. Ten of the 12 microsatellites used were composed of simple dinucleotide repeat motifs (Table 2). At these 10 loci, PCR products whose sizes differed by either two bases or multiples of two bases for each allele were elucidated. Of the two microsatellite loci with mixed repeat motifs, AVMIX04, gave PCR products whose sizes differed by two bases or multiples of two for each allele, whereas a battery of products, with varying sizes, were elucidated at the AVMIX03 locus. Test for departure from HW equilibrium (HWE) revealed deviation from HWE ( $P < 0.01$ ) in the entire Ghanaian population across only three (25%) of the 12 SSR loci with nine (75%) of the loci showing a population in HWE.

The allelic distributions, observed ( $H_o$ ) and expected ( $H_e$ ) heterozygosities, number of homozygotes, heterozygotes, and null alleles across all loci are summarized in Table 2. The 95 null alleles generated, representing 4.6% of all amplifications, ranged from five (AVAC01 and AVAG05) to 11 (AVAG06), averaging 7.9 across all loci. With the average  $H_o$  value (0.36) 29.4% less than  $H_e$  (0.51), the entire Ghanaian population was characterized by a state of heterozygote deficiency. Whereas the highest  $H_o$  (0.52) and  $H_e$  (0.77) were both recorded at the

Table 3. Distribution of allele sizes (basepairs) within all the 172 Ghanaian avocado seedlings analyzed by microsatellite markers in this study.<sup>z</sup>

Allele no.	Simple sequence repeat locus											
	AVAC01	AVAG03	AVMIX03	AVMIX04	AVAG05	AVAG06	AVAG07	AVAG10	AVAG11	AVAG13	AVAG22	AVAG25
1	111 (0.6)	107 (79.4)	145 (44.2)	168 (43.9)	95 (13.2)	64 (63.4)	108 (17.4)	175 (0.9)	109 (83.0)	104 (8.3)	112 (49.1)	104 (32.4)
2	113 (0.3)	109 (18.8)	147 (18.2)	178 (29.6)	99 (86.8)	70 (36.0)	110 (72.6)	177 (66.3)	113 (4.2)	106 (84.0)	116 (0.6)	110 (0.9)
3	115 (76.7)	113 (1.8)	151 (20.3)	180 (26.2)		72 (0.6)	112 (10.1)	179 (12.6)	123 (0.9)	126 (1.2)	122 (50.3)	120 (28.1)
4	121 (5.4)		153 (3.9)	192 (0.3)				191 (4.3)	133 (11.5)	128 (6.5)		122 (0.3)
5	123 (12.3)		155 (11.2)					195 (3.1)	135 (0.3)			140 (38.3)
6	125 (1.5)		166 (1.2)					197 (9.2)				
7	127 (3.3)		168 (0.9)					199 (3.7)				

<sup>z</sup>The allele frequencies (%) are indicated in parentheses.

AVMIX03 locus, the lowest  $H_o$  (0.20) was recorded at the locus AVAG13, whereas AVAG05 yielded the lowest  $H_e$  of 0.36. AVMIX03 yielded the least number of homozygote individuals (62) and the most (103) heterozygotes, whereas AVAG13 generated the most (142) homozygote and least (20) heterozygote individuals.

Nine alleles occurred at frequencies greater than 50%, of which six were also highly represented in the outlier populations, especially the West Indian race: 99 bp (AVAG05), 64 bp (AVAG06), 109 bp (AVAG11), 115 bp (AVAC01), 107 bp (AVAG03), and 110 bp (AVAG07) occurring at frequencies of 86.8%, 63.4%, 83.0%, 76.7%, 79.4%, and 72.6%, respectively.

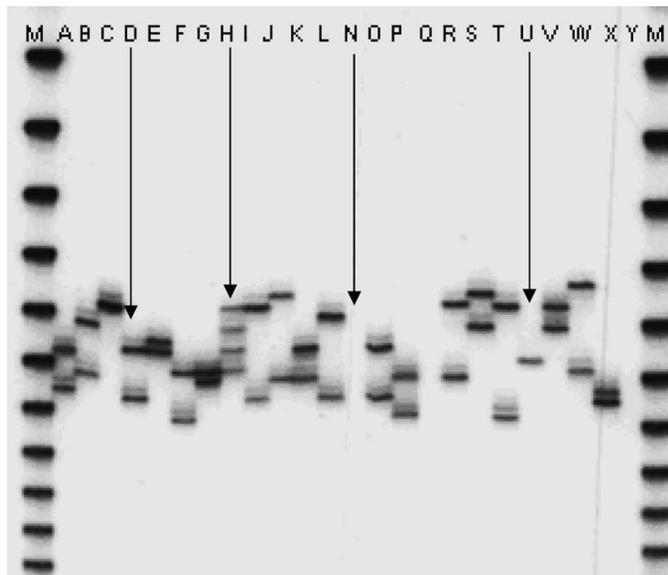


Fig. 2. Autoradiography of the various alleles generated by different Ghanaian avocado seedlings and cultivars at the AVAG25 locus. Lane M = DNA ladder (catalog no. 10821-015; Invitrogen, Carlsbad, CA), lane A = 'Aquila 1', lane B = 'Banos', lane C = 'Basaldua', lane D = 'Antigua', lane E = 'Galvan', lane F = 'Orizaba 1', lane G = 'Orizaba 3', lane H = 'Tantima', lane I = 'Amatenango', lane J = 'CH-C11 Padua', lane K = 'Comitan 1', lane L = 'Las Americas', lane N = 'IB Chalams', lane O = 'Puerto Viejo', lane P = 'Apakia 2', lane Q = 'Aquila 2', lane R = 'Nabal', lanes S–Y = seedlings from the Eastern region. Arrows point to a null allele (lane Q), multiple alleles (more than two alleles), and alleles of samples homozygote (lane U) and heterozygote (lane A) at AVAG25 locus. Multiple alleles were elucidated from 'Tantima' (lane H), whereas null alleles were elucidated from both 'IB Chalams' and 'Aquila 2' (lanes N and Q, respectively).

The first three alleles were completely absent in the Mexican population occurred at very low frequencies (less than 25%) in the Guatemalan population and at frequencies of 68.2%, 50.0%, and 54.6%, respectively, in the West Indian population. The fourth allele (115 bp at AVAC01) was absent in both the Guatemalan and Mexican populations but was 40.9% uniquely represented in the West Indian population. The allele 110 bp (AVAG07) was 59.1% represented in the West Indian population, 50.0% in the Mexican, and 37.5% in the Guatemalan population. The last allele occurred at frequencies of less than 20% in the Guatemalan and Mexican races but was 45.5% represented within the West Indian population.

Multiple (more than two) alleles were consistently elucidated at two loci (AVAG13 and AVAG25) in 'Tantima' (a West Indian cultivar). Three alleles, with sizes 106, 118, and 138 bp, were elucidated in the same individual at the AVAG13 locus, whereas four alleles (117, 121, 125, and 130 bp) were similarly generated at the AVAG25 locus (Fig. 2).

**LEVEL OF POLYMORPHISM WITHIN THE VARIOUS GHANAIAN SUBPOPULATIONS.** Assessing each regional subpopulation independently, an average of 2.5 alleles was generated by the Central population compared with an average of 0.9 recorded for the Brong-Ahafo population (Table 4). At the AVAG07 locus, all seedlings from the Central region were monomorphic, producing a 110-bp allele. Samples from the Volta region contained monomorphic allele (99 bp) at AVAG05 locus but were polymorphic across the other loci. All the other subpopulations were polymorphic across the 12 loci studied.  $H_e$  values of all regional subpopulations were higher than  $H_o$  values, indicating populations characterized by heterozygote deficiency. Volta region recorded the lowest  $H_o$  and  $H_e$  values of 0.28 and 0.45, respectively, whereas Brong-Ahafo and Ashanti regions recorded the highest  $H_o$  (0.42) and  $H_e$  (0.56) values, respectively. All the subpopulations had positive  $F_{IS}$  values (Table 4).

**GENETIC DISTANCES BETWEEN THE SUBPOPULATIONS AND PHYLOGENETIC TREES.** The genetic distances computed between the subpopulation pairs were very small ranging from 0.00 (between the Central-Eastern and Central-Volta subpopulation pairs) and 0.07 (between the Eastern-Brong-Ahafo and Volta-Brong-Ahafo subpopulation pairs). The genetic relationship among all 172 Ghanaian seedlings is represented in Figure 3A. All but 15 (8.7%) of the seedlings clustered together with high bootstrap values. These 'outlier samples' were of diverse regional backgrounds. The phylogenetic tree depicting the

Table 4. Characterization of the simple sequence repeat (SSR) loci among the various Ghanaian avocado subpopulations.

Population	N (no.) <sup>z</sup>	Mean alleles (no.)	Allele size range (basepairs)	A (%) <sup>y</sup>	$F_{IS}$ <sup>x</sup>	$H_o$ <sup>x</sup>	$H_e$ <sup>x</sup>
Eastern region	36	1.22	64–199	55.1	0.22	0.39	0.50
Central region	14	2.62	64–199	61.9	0.29	0.36	0.51
Volta region	13	0.91	64–197	67.3	0.38	0.28	0.45
Western region	32	1.00	64–199	59.1	0.34	0.35	0.53
Ashanti region	34	1.11	64–199	56.6	0.32	0.38	0.56
Brong-Ahafo region	43	2.71	64–197	55.8	0.14	0.42	0.49
Total	172						
Mean						0.36	0.51
SD						0.05	0.04

<sup>z</sup>N = number of individuals.

<sup>y</sup>A = percentage of homozygote individuals across all 12 SSR loci; homozygote (%) = 100% – [heterozygote (%) + null alleles (%)].

<sup>x</sup> $F_{IS}$  = inbreeding coefficient per sample overall loci;  $H_o$  = observed heterozygosity;  $H_e$  = unbiased expected heterozygosity (Levene, 1949).

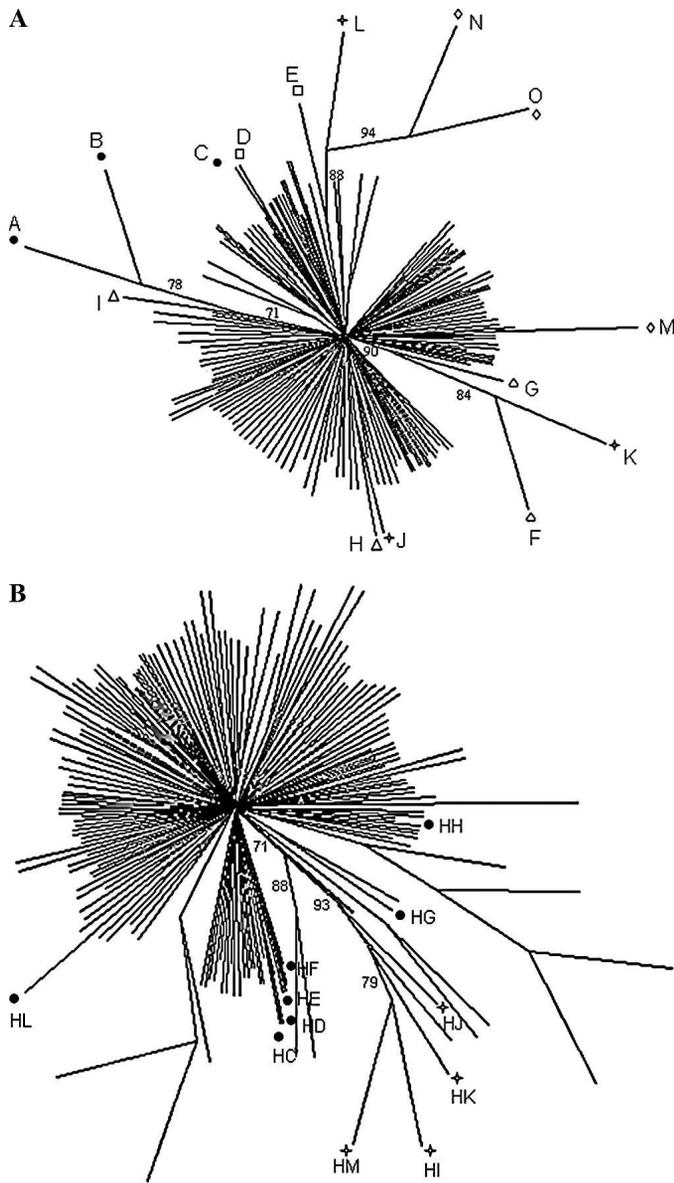


Fig. 3. Neighbor-joining (Saitou and Nei, 1987) phylogenetic trees based on Cavali-Sforza and Edwards (1967) chord distances calculated from the simple sequence repeat data. Bootstrap values were computed using 10,000 replications. Only bootstrap values greater than 50% are shown. (A) Tree of the 172 selected Ghanaian avocado seedling population. Samples that clustered separately are labeled A–O, and their regional distributions designated by the following shapes: ● = Eastern region samples, □ = Central region samples, △ = Volta region samples, △ = Western region samples, ✦ = Ashanti region samples, and ◇ = Brong-Ahafo region samples. (B) Tree of all 172 selected Ghanaian avocado seedling population (unlabeled) and representatives of West Indian race (designated by shapes and labels). ✦ West Indian cultivars that clustered outside the Ghanaian population: ‘Antigua’ (HI), ‘Galvan’ (HJ), ‘Orizaba 3’ (HK), and ‘Tantima’ (HM). ● West Indian, which clustered together with the Ghanaian population: ‘Bermercker’ (HC), ‘Orazaba 1’ (HD), ‘DW1 Bank’ (HE), ‘Tower 2’ (HF), ‘Butler’ (HG), ‘Fuchsia’ (HH), and ‘Waldin’ (HL).

genetic relatedness of the Ghanaian seedlings and West Indian representatives displayed an extra distinct cluster, which was supported by a high bootstrap value and consisted of four (36.4%) of the 11 West Indian representatives (Fig. 3B). These four West Indian cultivars are Antigua (HI), Galvan (HJ), Orizaba 3 (HK), and Tantima (HM). Consistent with the USAID reports on avocado introductions into Ghana, ‘Waldin’

and ‘Fuchsia’ clustered together with the Ghanaian seedlings. All the Mexican accessions genotyped clustered outside the Ghanaian seedlings, whereas all the Guatemalan accessions, except the cultivar Nabal, clustered outside the Ghanaian seedlings (figures not shown).

The genetic relationship among the Ghanaian samples (as a single population) and representatives of the three horticultural races is displayed in Figure 4. This phylogenetic tree grouped the three horticultural races into distinct clusters with high bootstrap values (99% and 100%). The Ghanaian population clustered with the West Indian race with a high (100%) bootstrap value, suggesting a higher genetic similarity between this race and the Ghanaian population.

### Discussion

Assessment of the genetic diversity of cultivated species and their wild relatives may aid in crop improvement strategies. Molecular markers such as SSRs are applied to assess diversity at the DNA level and thus provide an effective tool for selecting germplasm for conservation, the aim being to identify the minimum number of individuals representing most of the genetic variation. More saturated maps, using these DNA markers, have the potential to identify genes controlling important traits.

**CHARACTERISTICS OF SIMPLE SEQUENCE REPEATS AND POLYMORPHISM WITHIN THE GHANAIAN POPULATION.** SSRs are capable of detecting high levels of polymorphism, which are generated by either unequal crossing-over and/or replication slippages (Messier et al., 1996); however, the level of

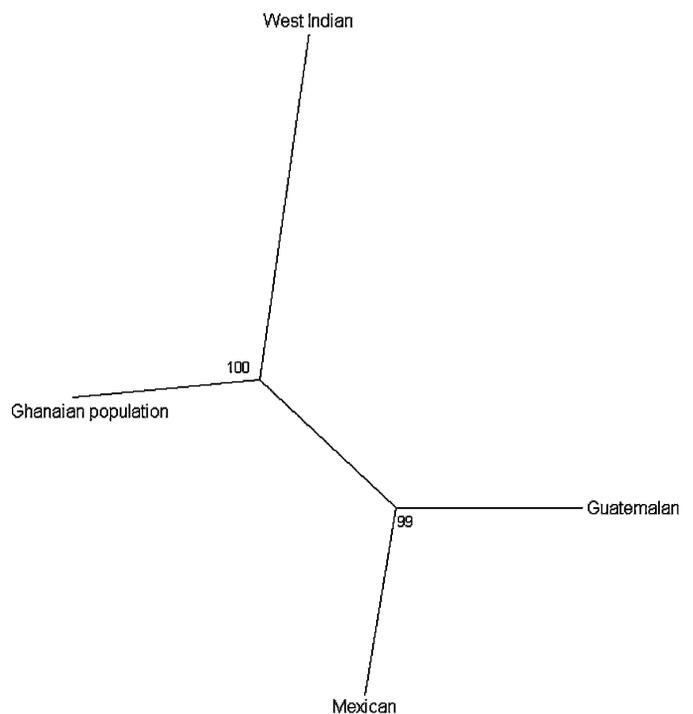


Fig. 4. Neighbor-joining (Saitou and Nei, 1987) tree of the entire Ghanaian population and the West Indian, Guatemalan, and Mexican races based on Cavali-Sforza and Edwards (1967) chord distances calculated from the simple sequence repeat data. Bootstrap values were computed using 10,000 replications.

polymorphism depends on the specific accessions or seedlings tested in each study. In this study, five of the 12 SSR loci had five or more alleles. Although AVAG05 was the least polymorphic locus in our population, Schnell et al. (2003), genotyping 258 accessions from the NGR (Miami, FL) and the University of California South Coast Research and Extension Center (Irvine, CA), identified AVAG07 as the least polymorphic. AVAC01, AVMIX03, and AVAG10 were the most polymorphic loci in our study, generating seven alleles each. Schnell et al. (2003), however, identified AVAG10 and AVAG13 as the most polymorphic. Whereas amplification products whose sizes differed by two or multiples of two were elucidated at four loci with dinucleotide repeat units (AVAG10, AVAG11, AVAG13, and AVAG21), Schnell et al. (2003) reported the generation of a battery of alleles of varying sizes at these loci. However, consistent with results reported by Schnell et al. (2003), we also identified a battery of products with varying sizes elucidated at the AVMIX03 locus. The 53 alleles elucidated from the 12 SSR loci ranged from two to seven per locus, whereas Schnell et al. (2003), using 14 SSR loci, identified 256 alleles ranging from eight to 30 per locus. The average number of alleles per locus was 4.4 in our study compared with 18.8 reported by Schnell et al. (2003). The significant differences between our results and that of Schnell et al. (2003) could be the result of the more genetically diverse accessions genotyped by the latter investigators. Using microsatellites, similar reports of high levels of polymorphism have also been reported in various species, including Ethiopian wheat (*Triticum aestivum* Jacubz.) (Teklu et al., 2006), grape (*Vitis vinifera* L.) (Martin et al., 2006), cowpea (*Vigna unguiculata* Walp.) (Diouf and Hilu, 2006), and rice (*Oryza sativa* L.) (Olufowote et al., 1997).

The allele with the highest frequency (86.8%) was recorded at AVAG05 (99 bp), whereas Schnell et al. (2003) identified the allele with 102 bp at the AVAG07 locus as the most frequent. The average gene diversity,  $H_o$  and  $H_e$ , reported in this study is lower than that detailed by Mhameed et al. (1996) and Schnell et al. (2003). Again, we attribute these differences to the higher genetic similarities among our Ghanaian seedlings.

Null alleles are attributed to technical errors during the amplification and/or mutations or aberrations within the primer binding sites. The mutations and aberrations may affect primer binding and hence the amplification (Gupta and Varshney, 2000; Liu et al., 2001). Because all DNA samples analyzed in this study were of good quality and quantity, and the null alleles were consistently and repeatedly obtained, the likelihood of a technical error for their generation is quite remote. We believe that sequence variation within the primer-binding sites of these samples could be the probable cause. Besides avocado (Borrone et al., 2007), null alleles have been reported in crops such as *T. aestivum* (Teklu et al., 2006).

The mean  $H_o$  (0.36) estimated across all loci, within the entire Ghanaian population, is lower than the  $H_e$  (0.51) and may be suggesting a state of heterozygote deficiency. Heterozygote deficiency usually results from nonrandom mating, inbreeding and, to some extent, limited gene flow among accessions (Chakraborty and Jin, 1992). In this study, however, the most likely cause of heterozygote deficiency could be inbreeding (including both selfing and crossing) among the genetically similar Ghanaian seedlings. The phenomenon is similar to extinction by hybridization (Allendorf et al., 2001; Ellstrand et al., 1999), where there is a rapid loss of genetic diversity

among individuals, which are genetically similar, resulting in a high number of homozygote individuals. This assertion is supported by the high percentage of homozygote individuals within the Ghanaian population and the NJ tree constructed for all 172 Ghanaian seedlings (Fig. 3A), suggesting very little genetic variation among the seedlings. This low genetic variation could be the result of either few founder parents or founder parents with very little genetic diversity.

The multiple alleles elucidated in the cultivar Tantina at AVAG13 (106/118 and 138 bp) and AVAG25 (117/121 and 125/130 bp) suggests duplication of parts of the chromosome(s) carrying the loci, resulting in the annealing of the SSR's primers to multiple sites. The phenomenon could also be as a result of nondisjunction of the SSR-carrying chromosome(s) during cell division (aneuploidy or trisomy). This theory is, however, unlikely because most cases of aneuploidy tend to have a major impact on the phenotype, which our morphological assessments of the cultivar did not verify. To establish the exact cause of the multiple alleles at these loci, the alleles elucidated should be mapped and sequenced.

**GENETIC VARIATIONS BETWEEN GHANAIAN SUBPOPULATIONS AND PHYLOGENETIC ANALYSES.** The various Ghanaian regional subpopulations were characterized by low genetic variation and also by a state of heterozygote deficiency expressed by  $H_o$  generally smaller than  $H_e$ . The low genetic variation is also depicted by the low genetic distance values determined between the subpopulation pairs. This population attribute, signifying low genetic divergence or differentiation and hence high level of inbreeding or genetic uniformity between the subpopulation pairs, is probably as a result of the fact that most of the Ghanaian avocado seedlings came from a common founder parent, the West Indian race, and are therefore genetically identical. These founder West Indian parents, providing most of the current planting materials, may have been too few or even if in large numbers and had very little genetic variability among them. Additionally, the ease of movement of plant materials from one region to another may have accounted for this low genetic variation. This assertion is supported by other results (A.K. Acheampong, R. Akromah, F.A. Ofori, and J.F. Takrama, unpublished data), in which farmers admitted using mostly seeds, obtained from unknown sources, as planting materials. Traditionally and historically, cocoa (*Theobroma cacao* L.), settler-farmers and inhabitants have been known to move frequently within the Eastern, Central, Western, Ashanti, and Brong-Ahafo regions as a result of cultural and language similarities. In fact, most of the cocoa farmers in the Western and Brong-Ahafo regions are known to have migrated from the Ashanti region. These farmers may have carried along some of these avocado seeds and planted them as intercrops on their new cocoa farms.

The presence of the other alleles such as 115 bp at AVAC01, 107 bp at AVAG03, 64 bp at AVAG06, and 106 bp at AVAG13 at relatively high frequencies within both the Ghanaian and West Indian populations is a strong indication of the genetic similarity between the Ghanaian and West Indian populations. The distribution of three alleles (64 bp at AVAG06, 177 bp at AVAG10, and 106 bp at AVAG13) exclusively found within the Ghanaian, West Indian, and Guatemalan populations is a further indication of genetic similarities between the two and also suggests introgression of some West Indian  $\times$  Guatemalan hybrids into Ghana. This lends more credence to the previously mentioned hypothesis that most, if not all, of the Ghanaian avocado founders were either West Indian cultivars or hybrids.

The phylogenetic trees (Figs. 3B and 4) strongly support the genetic characterization of the Ghanaian seedling population as belonging to the West Indian race and not the Mexican or Guatemalan, an assertion corroborated by morphological assessment of the Ghanaian seedlings made by coauthor A.K. Acheampong and supported by A. Ben-Ya'acov (personal communication). With their highly fissured barks, nail-head pedicels, shiny but leathery, pale green fruit skin, rough seed surfaces covered by thick seed coats, and the absence of anise scent in their leaves, the Ghanaian seedlings shared similar morphological traits with the West Indian race. None of the Mexican and Guatemalan cultivars (except Nabal) clustered with the Ghanaian samples, suggesting a large genetic distance between the Ghanaian population and the Mexican and Guatemalan races. This supports the notion that either no Mexican or Guatemalan cultivars were introduced into the country (i.e., all founders were of West Indian race) or those that were introduced did not survive the harsh tropical conditions of Ghana. 'Nabal' may have a specific genotype allowing it to survive the conditions in Ghana or it may have survived by mere chance. It is also possible that the SSRs we have used in this study were not sufficient to separate 'Nabal' from the Ghanaian seedlings and that additional SSRs could achieve this. The inadvertent "preferential selection" undertaken by the farmers may have favored West Indian accessions at the expense of the other two races. This may have resulted in genetic erosion of the Mexican race and, to some extent, the Guatemalan. It is worth noting that despite the small number of accessions analyzed, the three horticultural races were grouped into distinct clusters, a result similar to others (Davis et al., 1998; Mhameed et al., 1997; Schnell et al., 2003).

The results presented in this study demonstrate a low genetic variation between and within the Ghanaian avocado subpopulations. This study also reveals the high level of genetic similarity between the Ghanaian population and representatives of the West Indian race, which most probably constituted the bulk, if not all, of the founders. Regarding the avocado industry in Ghana, it is suggested that increasing Ghana's avocado gene pool by introducing various cultivars, including West Indian cultivars, hybrids, or both, will improve the Ghanaian avocado industry.

### Literature Cited

Allendorf, F.W., R.B. Leary, P. Spruell, and J.K. Wenberg. 2001. The problems with hybrids: Setting conservation guidelines. *Trends Ecol. Evol.* 16:102–108.

Ben-Ya'acov, A. 1995. The taxonomy of the avocado: A proposed new classification of the *Persea*, sub-Genus *Persea*. Proc. III World Avocado Congr., Tel Aviv, Israel. p. 3–7.

Borrone, J.W., R.J. Schnell, H.A. Violi, and R.C. Ploetz. 2007. Seventy microsatellite markers from *Persea americana* Miller (avocado) expressed sequence tags. *Mol. Ecol. Notes* 7:439–444.

Cavali-Sforza, L.L. and A.W.F. Edwards. 1967. Phylogenetic analysis: Models and estimation procedures. *Evolution Intl. J. Org. Evolution* 32:550–570.

Chakraborty, R. and L. Jin. 1992. Heterozygote deficiency, population substructure, and their implications in DNA fingerprinting. *Hum. Genet.* 88:267–272.

Chang, T.L., M.T. Lu, C.A. Lui, and L.Z. Chen. 2003. Genetic diversity analysis of Taiwan avocado accessions. *Actas V Congreso Mundial del Aguacate*, Malaga, Spain. p. 55–59.

Davis, J., D. Henderson, M. Kobayashi, and M.T. Clegg. 1998. Genealogical relationships among cultivated avocado as revealed through RFLP analysis. *J. Hered.* 89:319–323.

Diouf, D. and K.W. Hilu. 2006. Microsatellites and RAPD markers to study genetic relationships among cowpea breeding lines and local varieties in Senegal. *Genet. Resources Crop Evol.* 53:1305–1306.

Ellstrand, N.C., H.C. Prentice, and J.F. Hancock. 1999. Gene flow and introgression from domesticated plants into their wild relatives. *Annu. Rev. Ecol. Syst.* 30:539–563.

Felsenstein, J. 1989. PHYLIP-Phylogeny inference package (version 3.2). *Cladistics* 5:164–166.

Fiedler, J., G. Bufler, and F. Bangerth. 1998. Genetic relationships of avocado (*Persea americana* Mill.) using RAPD markers. *Euphytica* 10:249–255.

Ghana Division of Agriculture. 1961. Miscellaneous information. *Ghana Farmer* 8(1):2.

Gupta, P.K. and R.K. Varshney. 2000. The development and use of microsatellite markers for genetic analysis and plant breeding with emphasis on bread wheat. *Euphytica* 113:163–185.

Knight, R.J., Jr. 2002. History, distribution and uses, p. 1–14. In: Whitley, A.W., B. Schaffer, and B.N. Wolstenholme (eds.). *The avocado. Botany, production and uses*. CABI Publishing, Oxfordshire, UK.

Lavi, U., M. Akkaya, A. Bhagwat, E. Lahav, and P.B. Cregan. 1994. Methodology of generation and characteristics of simple sequence repeat DNA markers in avocado (*Persea americana* M.). *Euphytica* 80:171–177.

Lavi, U., J. Hillel, A. Vainstein, E. Lahav, and D. Sharon. 1991. Application of DNA fingerprint for the identification and genetic analysis of avocado. *J. Amer. Soc. Hort. Sci.* 116:1078–1081.

Levene, H. 1949. On a matching problem arising in genetics. *Ann. Math. Stat.* 20:91–94.

Liu, X.M., C.M. Smith, B.S. Gill, and V. Tolmay. 2001. Microsatellite markers linked to six Russian wheat aphid resistance genes in wheat. *Theor. Appl. Genet.* 102:504–510.

Martin, J.P., J.L. Santiago, O. Pinto-Carnide, F. Leal, M.C. Martinez, and J.M. Ortiz. 2006. Determination of relationships among autochthonous grapevine varieties (*Vitis vinifera* L.) in the northwest of the Iberian Peninsula by using microsatellite markers. *Genet. Resources Crop Evol.* 53:1255–1261.

Messier, W., S.H. Li, and C.B. Stewart. 1996. The birth of microsatellites. *Nature* 381:483.

Mhameed, S., J. Hillel, E. Lahav, D. Sharon, and U. Lavi. 1995. Genetic association between DNA fingerprint patterns and loci controlling agriculturally important traits in avocado. *Euphytica* 81:81–87.

Mhameed, S., D. Sharon, J. Hillel, E. Lahav, D. Kaufman, and U. Lavi. 1996. Level of heterozygosity and mode of inheritance of variable number of tandem repeat loci in avocado. *J. Amer. Soc. Hort. Sci.* 121:778–782.

Mhameed, S., D. Sharon, D. Kaufman, E. Lahav, J. Hillel, C. Degani, and U. Lavi. 1997. Genetic relationships within avocado (*Persea americana* Mill) cultivars and between *Persea* species. *Theor. Appl. Genet.* 94:279–286.

Miller, M.P. 1997. Tools for population genetic analysis (TFPGA) 1.3: A windows program for the analysis of allozyme and molecular population genetic data. 12 May 2008. <[http://www.marksgeneticssoftware.net/~vti\\_bin/shtml.exe/tfpga.htm](http://www.marksgeneticssoftware.net/~vti_bin/shtml.exe/tfpga.htm)>.

Morton, J.F. 1987. Avocado, p. 91–102. In: Morton, J.F. (ed.) *Fruits of warm climates*. Morton, Miami, FL.

Nei, M. 1978. Estimation of average heterozygosity and genetic distance from a small number of individuals. *Genetics* 89:583–590.

Olufowote, J.O., Y. Xu, X. Chen, W.D. Park, H.M. Beachell, R.H. Dilday, M. Goto, and S.R. McCouch. 1997. Comparative evaluation of within cultivar variation of rice (*Oryza sativa* L.) using microsatellite and RFLP markers. *Genome* 40:370–378.

Page, R.D.M. 1996. TREEVIEW: An application to display phylogenetic trees on personal computers. *Comput. Appl. Biosci.* 12:357–358.

- Popenoe, W. 1920. Manual of tropical and subtropical fruits. Mac-Millan, New York, NY.
- Popenoe, W. and G.A. Zentmyer. 1997. Early history of the avocado. California Avocado Soc. Yrbk. 81:163–171.
- Raymond, M. and F. Rousset. 2007. GENEPOP (version 3.4). 12 May 2008. <<http://genepop.curtin.edu.au/>>.
- Rhodes, A.M., S.E. Malo, C.W. Campbell, and S.G. Carmer. 1971. A numerical taxonomic study of the avocado (*Persea americana* Mill.). J. Amer. Soc. Hort. Sci. 96:391–395.
- Russell, J.R., F. Hosein, E. Johnson, R. Waugh, and W. Powell. 1992. Genetic differentiation of cocoa (*Theobroma cacao* L.) populations revealed by RAPD analysis. Mol. Ecol. 2:89–97.
- Saitou, N. and M. Nei. 1987. The neighbour-joining method: A new method for constructing phylogenetic trees. Mol. Biol. Evol. 4:406–425.
- Schnell, R.J., J.S. Brown, C.T. Olano, E.J. Power, and C.A. Krol. 2003. Evaluation of avocado germplasm using microsatellites markers. J. Amer. Hort. Sci. 128:881–889.
- Scora, R.W. and B.O. Bergh. 1990. The origins and taxonomy of avocado (*Persea americana* Mill.). Lauraceae. Acta Hort. 275:387–394.
- Sharon, D., P.B. Cregan, S. Mhameed, K. Kusharska, J. Hillel, E. Lahav, and U. Lavi. 1997. An integrated genetic linkage map of avocado. Theor. Appl. Genet. 95:911–921.
- Storey, W.B., B.O. Bergh, and G.O. Zentmyer. 1986. The origin, indigenous range and dissemination of the avocado. California Avocado Soc. Yrbk. 70:127–133.
- Stout, A.B. 1923. A study of the cross pollination of avocado in southern California. California Avocado Assn. Annu. Rpt. 1922–1923 8:29–45.
- Taah, K.J., P.G. Alderson, and J.B. Power. 2003. Molecular approaches for the characterization of Ghanaian avocado pear (*Persea americana* Mill.) germplasm. Actas V Congreso Mundial del Aguacate, Malaga, Spain. p. 69–72.
- Teklu, Y., K. Hammer, X.Q. Huang, and M.S. Roder. 2006. Analysis of microsatellite diversity in Ethiopian tetraploid wheat landraces. Genet. Resources Crop Evol. 53:1115–1126.
- Wright, S. 1978. Evolution and genetics of populations. Variability within and among natural populations. Vol. 4. Univ. Chicago Press, Chicago, IL.
- Yeh, F.C. and T.J.B. Boyle. 1997. Population genetic analysis of co-dominant and dominant markers and quantitative traits. Belg. J. Bot. 129:157 (abstr.).