J. Amer. Soc. Hort. Sci. 131(2):214–224. 2006.

# Mango Genetic Diversity Analysis and Pedigree Inferences for Florida Cultivars Using Microsatellite Markers

**R.J. Schnell[1], J.S. Brown, C.T. Olano, and A.W. Meerow**
*National Germplasm Repository, USDA, ARS, SHRS, 13601 Old Cutler Road, Miami, FL 33158*

**R.J. Campbell**
*Fairchild Tropical Botanic Garden, 10901 Old Cutler Road, Coral Gables, FL 33156*

**D.N. Kuhn**
*Department of Biological Sciences, Florida International University, Miami, FL 33199*

Additional index words. *Mangifera indica*, cultivar identification, SSR markers, mango breeding, parentage analysis

Abstract. **Mango (*Mangifera indica* L.) germplasm can be classified by origin with the primary groups being cultivars selected from the centers of diversity for the species, India and Southeast Asia, and those selected in Florida and other tropical and subtropical locations. Accessions have also been classified by horticultural type: cultivars that produce monoembryonic seed vs. cultivars that produce polyembryonic seed. In this study we used 25 microsatellite loci to estimate genetic diversity among 203 unique mangos (*M. indica*), two *M. griffithii* Hook. f., and three *M. odorata* Griff. accessions maintained at the National Germplasm Repository and by Fairchild Tropical Botanic Garden in Miami, Fla. The 25 microsatellite loci had an average of 6.96 alleles per locus and an average polymorphism information content (PIC) value of 0.552 for the *M. indica* population. The total propagation error in the collection (i.e., plants that had been incorrectly labeled or grafted) was estimated to be 6.13%. When compared by origin, the Florida cultivars were more closely related to Indian than to Southeast Asian cultivars. Unbiased gene diversity ($H_{nb}$) of 0.600 and 0.582 was found for Indian and Southeast Asian cultivars, respectively, and both were higher than $H_{nb}$ among Florida cultivars (0.538). When compared by horticultural type, $H_{nb}$ was higher among the polyembryonic types (0.596) than in the monoembryonic types (0.571). Parentage analysis of the Florida cultivars was accomplished using a multistage process based on introduction dates of cultivars into Florida and selection dates of Florida cultivars. In total, 64 Florida cultivars were evaluated over four generations. Microsatellite marker evidence suggests that as few as four Indian cultivars, and the land race known as 'Turpentine', were involved in the early cultivar selections. Florida may not represent a secondary center of diversity; however, the Florida group is a unique set of cultivars selected under similar conditions offering production stability in a wide range of environments.**

Mango is a traditional, highly esteemed crop in India and Southeast Asia. Over the last 500 years, it has become well established in tropical American locations, including Florida and Hawaii in the United States (Popenoe, 1920). The introduction of mangos into Florida and subsequent development of a Florida group of mangos has been reviewed by Knight and Schnell (1994), and Florida has been considered a secondary center of genetic diversity for the species. Mango cultivars have also been classified based on the type of embryo developed: monoembryonic where a single zygotic embryo produces a single shoot and polyembryonic where nucellar embryos also develop and multiple shoots are produced at germination.

The Florida mango cultivars are unique in that they are hybrids between Indian cultivars (primarily monoembryonic) and the Southeast Asian cultivars (primarily polyembryonic) selected under southern Florida conditions. These Florida selections are widely grown commercial cultivars affording production stability across many environments (Mukherjee, 1997). Understanding genetic relationships among the Florida cultivars and their relationship to both the Indian and Southeast Asian races is important for identification of genes involved with wide adaptation and for the efficient development of improved cultivars.

Isozymes were the first markers used to fingerprint mango cultivars, to determine self vs. cross pollination, and to estimate genetic relationships (Degani et al., 1990; Knight and Schnell, 1994). Randomly amplified polymorphic DNA (RAPD) markers were also used to fingerprint cultivars and estimate genetic relationships in mango (Schnell et al., 1995). A group of 'Haden' seedlings and a random group of seedlings were evaluated using 11 RAPD primers. This study supported the 'Haden' parentage of 'Eldon', 'Lippens', 'Tommy Atkins', and 'Zill'; however, the parentage of 'Glenn' and 'Osteen' was questioned. Adato et al. (1995) used DNA fingerprinting (DFP) to evaluate genetic relationships between 26 mango cultivars and 14 mango rootstocks. They provided a pedigree that further confirmed the relationship between many of the 'Haden' seedlings. Lopez-Valenzuela et al. (1997) used RAPD markers to estimate genetic diversity among 15 mango rootstock cultivars using 13 markers. They identified a specific RAPD band associated only with the polyembryonic types. Eiadthong et al. (1999) utilized anchored simple-sequence repeat markers to analyze 22 mango cultivars. They were able to distinguish genotypes, but were unable to find markers unique to either monoembryonic or polyembryonic types, or for the Thai cultivars selected for green harvest (crispy mango) from the cultivars selected for ripe fruit production. Kashkush et al. (2001) utilized amplified fragment-length polymorphisms (AFLP) to estimate genetic relationships between 16 cultivars and seven rootstock cultivars. They also analyzed 29 progeny from a cross of 'Tommy-Atkins' and 'Keitt' and produced a crude linkage map that identified 13 of the 20 linkage groups.

Received for publication 24 June 2005. Accepted for publication 16 Oct. 2005.
[1]To whom reprint requests should be addressed. E-mail address: rschnell@saa.ars.usda.gov

Viruel et al. (2005) developed the first reported set of 16 microsatellite markers for mango, of which 14 produced the expected one or two amplification products per genotype. These 14 microsatellites were used to evaluate 28 mango genotypes that included 14 Florida cultivars. Discrimination of all 28 genotypes was possible and the average number of alleles per locus was 5.3. Previously known pedigree information for the 'Haden' family of mangos was confirmed and was in agreement with previously published RAPD and DFP analyses (Adato et al., 1995; Schnell et al., 1995) with one exception; Viruel's clone of 'Zill' was not resolved as a seedling of 'Haden'. Schnell et al. (2005) developed a second set of 15 microsatellite markers and analyzed 59 Florida cultivars and four related species. Two of the microsatellites were monomorphic among the Florida cultivars; the other 13 had an average number of alleles per locus of 4.2 with PIC values varying from 0.21 to 0.63.

Mango was introduced to Florida in the 19th century. The earliest introductions were from the West Indies and India, followed by the introduction of several hundred accessions in the 20th century from Southeast Asia, India, and from other mango growing areas of the world [Florida Mango Forum (FMF), 1951]. Florida selections are not the result of formal breeding programs. Early Florida selections were made by growers and enthusiasts and historical information is often anecdotal. The FMF, established in 1938 for the advancement of mango production, documented historical information on the parentage of Florida cultivars in their proceedings. In addition to the USDA Germplasm Resources Information Network (GRIN) database, several sources compile information on Florida mango selections and introduction of accessions to Florida (Campbell, 1992; Ruehle and Ledin, 1955; Singh, 1960). These sources were used to determine when Florida cultivars first fruited, were selected, and possibly contributed to future cultivar development. Introduced cultivars were also segregated by date of introduction to estimate their contribution to the development of the Florida cultivars.

Our objectives were to estimate genetic diversity in a large collection of mango maintained at the USDA National Germplasm Repository (NGR) and Fairchild Tropical Botanic Garden (FTBG) in Miami, Fla., using microsatellite markers. Microsatellites have advantages over other types of molecular markers, i.e., their abundance in most genomes, uniform distribution, hypervariability, codominance, and PCR-based protocols (Li et al., 2002). We wanted to characterize and contrast the Indian (primarily monoembryonic) cultivars, the Southeast Asian (primarily polyembryonic) cultivars, and the Florida cultivars. Since all plants (clones) of a given cultivar were genotyped, we also investigated the fidelity of germplasm propagation within cultivars, as clonal collections are known to contain identical genotypes with different names and mixtures of genotypes with the same name (Schnell et al., 1999). Confirmation of reported parents and identification of unknown parents of the Florida cultivars was also a major objective. Parental identification would enhance current cultivar development and provide a basis for future genetic studies.

## Materials and Methods

**Plant materials.** Leaf material was sampled from 359 plants in the *Mangifera* L. germplasm collection at the NGR in Miami, Fla. Cultivars with a single representative plant totaled 137. The remaining 72 cultivars were represented by multiple plants with a total of 222. These 222 plants include 38 cultivars with two, 15 cultivars with three, 11 cultivars with four, two cultivars with five, two cultivars with six, two cultivars with seven, one cultivar with eight, and one cultivar with 13 plants. The backgrounds of the cultivars are listed in Table 1. The plant material is broadly categorized into groups by geographic origin. The populations (with over four cultivars) included: Florida (64), Southeast Asia (34), India (29), the West Indies (26), Africa (16), Hawaii (9), Central America (7), Israel (6), South America (6), and Pacific (4). The cultivar Siamese, of unknown origin, and the cultivar Edgehill, selected in California, were not placed in these groupings. The Southeast Asian group includes: Borneo, Burma, Indonesia, Philippines, Vietnam, Australia, Indo-China, and Thailand. Plants originating from Cuba, Puerto Rico, Haiti, Jamaica, Trinidad, and the West Indies are considered West Indian. The out-group consisted of two accessions of *M. griffithii*, and three accessions of *M. odorata* (Table 1). DNA extraction was performed on leaf tissue as described by Schnell et al. (2005).

**Microsatellite markers and polymerase chain reaction (PCR) amplification.** Twelve of the microsatellite markers used in this study were previously reported by Viruel et al. (2005), and one by Duval et al. (2005), and 12 were developed at the NRG-Miami (Schnell et al., 2005). Forward primers were labeled with a fluorescent dye on the 5′ end and all 25 primer pairs were used on all individuals for the analysis. Microsatellite loci names are listed in Table 2. PCR amplification reactions were carried out as described by Schnell et al. (2005) on a DNA Engine tetrad thermal cycler (MJ Research, Watertown, Mass.).

**Electrophoresis.** Capillary electrophoresis was performed on either an ABI Prism 3100 Genetic Analyzer or an ABI Prism 3730 Genetic Analyzer (Applied Biosystems, Foster City, Calif.) as described by Schnell et al. (2005). Resulting data were analyzed with GeneMapper 3.0 (Applied Biosystems) for internal standard and fragment size determination and for allelic designations.

**Cultivar identification.** The microsatellite genotype of each plant was compared against every other genotype in the dataset in order to find unique genotypes for each cultivar. Single plant cultivars that had matching genotypes were considered synonymous. If all samples for multiple plant cultivars were not identical, then the most common genotype was used to represent that cultivar; nonmatching genotypes were considered propagation errors. The addition of microsatellite fingerprints to accession information is now standard practice for germplasm management in the National Plant Germplasm System (NPGS).

**Diversity analysis.** Polymorphism information content values for each locus were calculated as follows:

$$PIC = 1 - \sum \hat{p}_{ij}^{2}$$

where $p_{ij}$ is the frequency of the $j$th allele for marker $i$, and summation extends over n alleles (Liu, 1998).

Unbiased gene diversity ($H_{nb}$) and observed heterozygosity ($H_{obs}$; Nei, 1987) were estimated from the allele frequencies of the monoembryonic and polyembryonic types as well as from the 10 populations based on geographical distribution using GENETIX (version. 4.0; Université de Montpellier II, Montpellier, France) (Table 3). The relationships between the populations were represented using a dendrogram constructed from allele frequencies averaged over populations using modified Rogers' distance (Wright, 1978) and the unweighted pair group method using arithmetic average (UPGMA). Bootstrapping was accomplished using 1000 replications with the program Populations (Langella, 2002).

Table 1. Listing by origin of the 208 *Mangifera* spp. cultivars and accessions evaluated with 25 microsatellite markers. Horticultural type designated as monoembryonic (M), polyembryonic (P), or unknown (U). Parent–Offspring sets for parentage analysis of four generations of Florida selections: I (1880–1910), II (1910–40), III (1940–60), IV (1960–present).

| Cultivar | Origin | Type | Set |
|---|---|---|---|
| Tyler Premier | Africa | M | IV |
| Sabre | Africa, East | P | IV |
| Heidi | Africa, South | M | IV |
| Nelpetite | Africa, South | M | IV |
| Piva | Africa, South | M | IV |
| Kensington | Australia | P | IV |
| *M. griffithii* | Borneo | U | |
| *M. griffithii*/*M. indica* | Borneo | U | |
| *M. odorata* (93-890) | Borneo | U | |
| *M. odorata* 'Rampangi' | Borneo | U | |
| *M. odorata*(A)/*M. indica* | Borneo | U | |
| Tenom | Borneo | P | IV |
| Extrema (R8-T3) | Brazil | P | IV |
| Itamaraca | Brazil | M | I |
| Mendonsa | Brazil | P | IV |
| Rosa | Brazil | P | IV |
| Aung Dim | Burma | P | IV |
| Burma (Myatryaut) | Burma | P | IV |
| Pam Kai Mia | Burma | P | IV |
| Po Piju Kalay (PPK) | Burma | P | IV |
| Seinta Lone | Burma | U | IV |
| Shwehintha (Swethintha) | Burma | P | IV |
| Edgehill | California | M | IV |
| Vallenato | Colombia | P | IV |
| Chino | Cuba | P | II |
| Guanabacoa Late | Cuba | U | IV |
| Prieto | Cuba | U | IV |
| Toledo | Cuba | P | IV |
| Bullock's Heart | Egypt | P | IV |
| Diab | Egypt | P | IV |
| Egypt | Egypt | U | IV |
| Ewais | Egypt | P | IV |
| Gaylour | Egypt | M | IV |
| Hindi Besennari | Egypt | M | IV |
| Hort 1 | Egypt | P | IV |
| Zebda | Egypt | M | IV |
| Fiji Long | Fiji | P | IV |
| Fiji Short | Fiji | P | IV |
| Allen-King / Everbearing | Florida | U | IV |
| Anderson | Florida | M | II |
| Baileys Marvel | Florida | M | IV |
| Becky | Florida | M | IV |
| Becky FF | Florida | U | IV |
| Beverly | Florida | M | IV |
| Brooks | Florida | M | I |
| Carrie | Florida | M | III |
| Cogshall | Florida | M | III |
| Cushman | Florida | M | II |
| Dot | Florida | M | IV |
| Duncan | Florida | M | IV |
| Dupuise | Florida | M | IV |
| Earlygold | Florida | M | III |
| Edward | Florida | M | II |
| Eldon | Florida | M | III |
| Eulogio | Florida | M | IV |
| Florigon sdlg | Florida | P | IV |
| Florigon | Florida | P | II |
| Ford | Florida | M | IV |
| Gary | Florida | P | IV |
| Glenn | Florida | M | III |
| Golden Lippens | Florida | M | III |
| Golden Nugget | Florida | M | IV |
| Gootee | Florida | U | IV |
| Haden | Florida | M | I |
| Hatcher | Florida | M | IV |
| Hodson | Florida | M | III |
| Iris | Florida | M | IV |
| Irwin | Florida | M | III |
| Jacquelin | Florida | M | III |
| Jakarta | Florida | M | IV |
| Jewel | Florida | M | IV |
| Joellen | Florida | P | IV |
| Jubilee | Florida | M | IV |
| Keitt (Jo Z) | Florida | M | II |
| Kent | Florida | M | II |
| Lathrop | Florida | U | IV |
| Lily | Florida | M | IV |
| Lippens | Florida | M | II |
| Mapulehu | Florida | M | IV |
| Martin | Florida | M | IV |
| Merrit Island | Florida | U | IV |
| Osteen | Florida | U | III |
| Palmer | Florida | M | II |
| Pettigrew | Florida | M | III |
| Rosigold | Florida | P | IV |
| Ruby | Florida | M | III |
| S-01 | Florida | M | II |
| S-19 | Florida | U | II |
| Saicra | Florida | P | IV |
| Sensation | Florida | M | III |
| Southern Blush | Florida | M | IV |
| Spirit of 76 | Florida | M | IV |
| Springfels | Florida | M | II |
| Sunset | Florida | M | III |
| Tommy Atkins | Florida | M | III |
| Torbet | Florida | M | IV |
| Valencia Pride | Florida | M | III |
| Van Dyke | Florida | M | II |
| Winters | Florida | P | IV |
| Z-80 | Florida | M | IV |
| Ziiate | Florida | M | IV |
| Zill | Florida | M | II |
| Buxton Spice | Guyana | P | IV |
| Madame Francis | Haiti | P | II |
| CAC | Hawaii | U | IV |
| Excel | Hawaii | U | IV |
| Fukuda | Hawaii | M | IV |
| Ono | Hawaii | P | II |
| Paris Selection No. 1 | Hawaii | P | IV |
| Rapoza | Hawaii | M | IV |
| St. Maui | Hawaii | M | IV |
| Shindiri | Hawaii | M | IV |
| Step | Hawaii | U | IV |
| Lancetilla | Honduras | M | IV |
| Alampur Banishan (Imann Pasand) | India | M | IV |
| Alfonso | India | M | I |
| Alphonse (White Alfonso) | India | M | I |
| Ameeri | India | M | I |
| Amin Abrahimpur | India | U | II |
| Amini | India | M | I |
| Bombay (Paheri) | India | M | I |
| Borsha (Borsha of Bhadgaon) | India | M | I |
| Cowasju Patel | India | M | I |
| Cowasjee Patel x Piri | India | U | II |
| Fernandez | India | M | I |
| Himsagar (Fajri) | India | M | I |
| Janardhan Pasand | India | M | IV |
| Jehangir | India | M | IV |
| Langra | India | M | I |
| Madras | India | M | I |
| Mahmoods Vikarabad | India | M | IV |
| Malda | India | U | I |
| Mallika | India | M | IV |
| Mulgoa | India | M | II |
| Mulgoba | India | M | I |
| Neelum | India | M | II |
| Pachmari Hills | India | U | III |
| Panchadarakalasa | India | M | IV |
| Pedda Rasam | India | M | IV |
| Royal Special | India | M | IV |
| Rumani | India | U | II |
| Sandersha | India | M | I |
| Vanraj | India | M | IV |
| Saigon1 | Indo-China | M | I |
| Saigon3 | Indo-China | M | I |
| Saigon4 | Indo-China | M | I |
| Saigon5 | Indo-China | M | I |
| Saigon6 | Indo-China | M | I |
| Saigon7 | Indo-China | M | I |
| Aroemanis (Golek) | Indonesia | P | II |
| Madoe | Indonesia | P | II |
| 13-01 | Israel | P | IV |
| Leshem | Israel | M | IV |
| Magshimim | Israel | M | IV |
| Maya | Israel | M | IV |
| Naomi | Israel | M | IV |
| Tahar | Israel | M | IV |
| East Indian | Jamiaca | M | IV |
| Long | Jamiaca | U | IV |
| Number 11 | Jamiaca | P | I |
| White Pirie | Jamiaca | M | IV |
| Malindi 1 | Kenya | P | IV |
| Malindi 2 | Kenya | P | IV |
| Ngowe | Kenya | P | IV |
| Diplomatico | Mexico | M | IV |
| Manila | Mexico | P | I |
| Manilita | Mexico | P | IV |
| Manzanillo | Mexico | M | IV |
| Oro | Mexico | M | IV |
| Tuehau | Pacific | U | II |
| Tutehau | Pacific | U | II |
| Fairchild | Panama | M | II |
| Carabao (Phillipine) | Philippines | P | I |
| Mamita (Merrit Island Saigon) | Puerto Rico | P | IV |
| Manzana Colorada | Puerto Rico | P | IV |
| Tete Nene | Puerto Rico | P | II |
| Braham Kei Meu | Thailand | P | IV |
| Hong Sa | Thailand | P | IV |
| Ivory | Thailand | P | IV |
| Katar Rum Rung (Ratar Rum Rung) | Thailand | P | IV |
| Kyo Savoy | Thailand | P | IV |
| Mamnou (Mamou, Praya Sowoy) | Thailand | M | IV |
| Nam Doc Mai | Thailand | P | IV |
| Nam Tan Teen (Nam Tam Teem) | Thailand | P | IV |
| Okrung (Okrung Tong) | Thailand | P | IV |
| Pohn Sawadee | Thailand | M | IV |
| Rataul (Phimsen Mun) | Thailand | P | IV |
| Sad Lium Pua | Thailand | P | IV |
| Sang Tong | Thailand | P | IV |
| Suwon Tip | Thailand | P | IV |
| Thai Everbearer | Thailand | P | IV |
| Tong Dani | Thailand | P | IV |
| Graham | Trinidad | M | II |
| Ice Cream | Trinidad | M | IV |
| Tobago Small Red | Trinidad | P | IV |
| Siamese | unknown | P | IV |
| Cambodiana | Vietnam | P | I |
| Colombo Kidney | West Indies | U | II |
| Julie | West Indies | M | I |
| Peach | West Indies | P | I |
| Pere Louis | West Indies | U | I |
| Turpentine1 | West Indies | P | I |
| Turpentine10 | West Indies | P | I |
| Turpentine11 | West Indies | P | I |
| Turpentine3 | West Indies | P | I |
| Turpentine7 | West Indies | P | I |
| Turpentine8 | West Indies | P | I |
| Turpentine9 | West Indies | P | I |

Table 2. Microsatellite loci used in the analysis of the *Mangifera indica* germplasm collections; $H_E$ = expected heterozygosity, $H_O$ = observed heterozygosity, PIC = polymorphic information content, HWE = Hardy–Weinberg equilibrium.

| Locus | GeneBank accession no. | Alleles (no.) | Size range (bp) | $H_O$ | $H_E$ | PIC values | Departure from HWE |
|---|---|---|---|---|---|---|---|
| MiSHRS-1 | AY942817 | 7 | 191–209 | 0.736 | 0.773 | 0.736 | ** |
| MiSHRS-4 | AY942818 | 6 | 121–133 | 0.668 | 0.669 | 0.622 | |
| MiSHRS-18 | AY942819 | 8 | 90–117 | 0.602 | 0.682 | 0.624 | |
| MiSHRS-26 | AY942821 | 3 | 260–275 | 0.260 | 0.247 | 0.220 | |
| MiSHRS-29 | AY942822 | 5 | 174–184 | 0.483 | 0.493 | 0.448 | |
| MiSHRS-30 | AY942823 | 4 | 221–232 | 0.046 | 0.075 | 0.073 | ** |
| MiSHRS-32 | AY942824 | 12 | 200–226 | 0.416 | 0.599 | 0.557 | ** |
| MiSHRS-33 | AY942825 | 6 | 236–257 | 0.366 | 0.380 | 0.354 | |
| MiSHRS-34 | AY942826 | 2 | 228–231 | 0.005 | 0.005 | 0.005 | |
| MiSHRS-37 | AY942828 | 7 | 125–137 | 0.678 | 0.663 | 0.611 | |
| MiSHRS-39 | AY942829 | 9 | 345–369 | 0.682 | 0.656 | 0.590 | |
| MiSHRS-44 | AY942830 | 7 | 245–278 | 0.158 | 0.484 | 0.400 | ** |
| LMMA1 | AY628373 | 10 | 195–215 | 0.675 | 0.824 | 0.798 | ** |
| LMMA4 | AY628376 | 6 | 222–244 | 0.632 | 0.688 | 0.624 | |
| LMMA5 | AY628377 | 3 | 278–282 | 0.305 | 0.314 | 0.285 | |
| LMMA7 | AY628379 | 7 | 198–214 | 0.753 | 0.767 | 0.734 | |
| LMMA8 | AY628380 | 9 | 254–270 | 0.782 | 0.776 | 0.743 | |
| LMMA9 | AY628381 | 7 | 171–187 | 0.677 | 0.833 | 0.808 | ** |
| LMMA10 | AY628382 | 12 | 150–188 | 0.826 | 0.800 | 0.775 | ** |
| LMMA11 | AY628383 | 9 | 230–248 | 0.769 | 0.787 | 0.754 | |
| LMMA12 | AY628384 | 7 | 198–206 | 0.711 | 0.743 | 0.707 | |
| LMMA14 | AY628386 | 5 | 162–171 | 0.479 | 0.528 | 0.444 | |
| LMMA15 | AY628387 | 7 | 207–221 | 0.616 | 0.625 | 0.576 | ** |
| LMMA16 | AY628388 | 8 | 211–242 | 0.721 | 0.761 | 0.720 | ** |
| mMiCIR014 | AJ635176 | 8 | 148–165 | 0.570 | 0.640 | 0.601 | ** |

**Significant at $P \leq 0.01$.

The principal coordinate analyses [PCA (Gower, 1966)] were performed using SAS (version 9.0 for Windows; SAS Institute, Cary, N.C.) with modified Rogers' distance for the Florida, India, West Indian, and Southeast Asian populations. PCA is a scaling or ordination method that uses eigenvalues and eigenvectors derived from a distance matrix that has been "double centered." Principal coordinate scores are then produced for each original observation, one value being obtained for each axis. The final step is to plot the coordinate scores in a two- or three-dimensional plot.

**Parentage analysis.** For the Florida cultivars with known and unknown parents, parentage analyses were performed in a multistep process by chronological stages using four sets of candidate parents and offspring (Table 1). Introduced cultivars known to be in Florida when the Florida cultivars were selected were included in the analyses, as well as all Florida cultivars in the collection. The first set includes 35 cultivars introduced between 1880 and 1910, and the first Florida selections, 'Haden' and 'Brooks'. The second set included the original 35, 'Haden', 'Brooks', and 16

Table 3. Allele sizes estimated from the ABI 3730 for the 25 microsatellite loci for all *Mangifera* spp. analyzed. Alleles unique to *M. odorata*, *M. griffithii*, and *M. indica* population groups by origin are indicated as follows: a = Africa, h = Hawaii, i = India, p = Pacific, sa = Southeast Asia, g = *M. griffithii*, and o = *M. odorata*. Average frequency is indicated in parenthesis for alleles occurring at high frequency in all populations.

| Locus | Alleles | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| MiSHRS1 | 191 | 197 | 199 | 201 | 205 | 207 | 209 | 213[o] | | | | |
| MiSHRS4 | 121 | 125 | 127 | 129 | 131 | 133 | 141[g] | 151[o] | | | | |
| MiSHRS18 | 90 | 96 | 99 | 102 | 105 | 111 | 114[sa] | 117[i] | | | | |
| MiSHRS26 | 260 | 263[o] | 269[g] | 272(0.862) | 275 | | | | | | | |
| MiSHRS29 | 172[o] | 174 | 178(0.679) | 180 | 182 | 184[i] | | | | | | |
| MiSHRS30 | 221 | 222(0.960) | 229 | 232[i] | | | | | | | | |
| MiSHRS32 | 200 | 201 | 202[sa] | 204 | 206 | 208 | 210 | 214 | 216[a] | 218 | 224 | 226 |
| MiSHRS33 | 236 | 239 | 242[g] | 245(0.775) | 248 | 251 | 257 | | | | | |
| MiSHRS34 | 228(0.997) | 231[i] | 234 | 240[o] | | | | | | | | |
| MiSHRS37 | 120[g] | 125 | 127 | 129 | 131 | 132 | 135[i] | 137 | | | | |
| MiSHRS39 | 345[a] | 348 | 351 | 354[p] | 357 | 360 | 363[a] | 366 | 369 | 372[g] | 384[g] | |
| MISHRS44 | 245[sa] | 251 | 253 | 258 | 260(0.642) | 276[h] | 278[h] | | | | | |
| LMMA1 | 193[o] | 195 | 197 | 199 | 201 | 203 | 205 | 207 | 209[i] | 211 | 215[i] | |
| LMMA4 | 222 | 228 | 229[g] | 230[o] | 238 | 240[i] | 242 | 244 | | | | |
| LMMA5 | 278(0.813) | 280 | 282 | 284[g] | | | | | | | | |
| LMMA7 | 198 | 200 | 202[g] | 204 | 206 | 208 | 212 | 214 | 216[o] | 220[o] | | |
| LMMA8 | 254 | 258 | 260 | 262[sa] | 264 | 265 | 266 | 268 | 270 | 271[o] | 279[o] | |
| LMMA9 | 171 | 177 | 179 | 181 | 183 | 185 | 187 | | | | | |
| LMMA10 | 150 | 152 | 154[i] | 156 | 158 | 162 | 170 | 172[i] | 174 | 178 | 180 | 188 |
| LMMA11 | 222[o] | 224[o] | 230 | 234[i] | 236 | 238 | 240 | 242 | 244 | 246[a] | 248 | |
| LMMA12 | 198 | 199 | 200 | 201 | 202 | 204 | 206 | | | | | |
| LMMA14 | 162 | 163 | 165 | 169 | 171 | | | | | | | |
| LMMA15 | 207 | 209 | 211 | 213 | 217 | 219 | 221 | 225[g] | | | | |
| LMMA16 | 211[sa] | 231 | 233[g] | 234[a] | 236 | 239 | 240[i] | 241 | 242 | 243[o] | | |
| mMiCIR14 | 148 | 153 | 154 | 156 | 158 | 159 | 160[g] | 162 | 165[i] | 167[o] | | |

introductions made between 1910 and 1940. These were used to estimate the parentage of 13 Florida cultivars selected between 1910 and 1940. The third set included the previous 66 (Set I and II), and one new candidate parent. These were used to estimate the parentage of 16 cultivars selected in Florida between 1940 and 1960. The final set (Set IV) included the previous 83 cultivars as potential parents (Sets I, II, and III) along with 87 additional accessions imported into Florida between 1948 and 1973. These were included as potential parents for 33 accessions selected in Florida from 1960 to 2000. Parentage analysis was performed using the program CERVUS (Marshall et al., 1998; Slate et al., 2000). This software uses a simulation program to generate log-likelihood scores and provides a confidence statistic for assigning paternity. Simulations were accomplished using 50,000 replications with no genotyping errors allowed. The analysis was conducted for Sets I, II, and III using three steps. First, the most-likely parent was identified and then made the known parent followed by another estimation of the second most-likely parent. Second, the analysis was run asking for all possible parents with positive sum of the log-likelihood ratios at each locus (LOD scores). Third, to investigate unresolved pedigrees, genotyping error rate of 0.01% was allowed and the analysis re-run. For Set IV an extra step was performed where all potential parents with negative LOD scores based on the first analysis were deleted and the analysis re-run. Negative LOD scores indicate that the candidate parent is less likely than a parent selected at random from the population. Inbreeding among the Florida cultivars was estimated using the Proc Inbreed procedure of SAS (version 9.0 for Windows).

## Results

**Levels of polymorphism.** The 25 microsatellite loci were highly variable among the *M. indica* populations. The number of alleles varied from two (MiSHRS34) to 12 (LMMA10 and MiSHRS32), with an average of 6.96 alleles per locus and average PIC value of 0.552 (0.005–0.808; Table 2). Ten loci departed significantly from Hardy–Weinberg equilibrium [HWE ($P < 0.01$)]. Heterozygote deficiency ($P > 0.01$) was detected at all but one locus (LMMA10) not in HWE. At locus MiSSR34 a single heterozygote was detected (Table 2).

Five of the microsatellites were composed of trinucleotide repeats and all alleles differed by multiples of three bases. Nineteen of the microsatellites were composed of dinucleotide repeats and most alleles differed by multiples of two bases. Alleles with single base length differences were found for locus MiSHRS37, MiSHRS30, MiSHRS32, LMMA4, LMMA8, LMMA12, LMMA14, LMMA16, and mMiCIR14. MiSHRS30 is a seven nucleotide repeat with two of the four alleles not differing by a complete repeat unit (data not shown). MiSHRS44 is a complex dinucleotide repeat with alleles detected that do not differ by complete repeat units (Table 3).

**Cultivar identification.** Among the 72 cultivars with multiple plants three were found to be mixtures of genotypes. Thirteen plants of 'Turpentine', the common Florida rootstock, exist on the NGR-Miami and consist of seven related genotypes. Of these, one genotype ('Turpentine 1') was found in five plants, while three plants of a second genotype ('Turpentine 11') exist. The other five 'Turpentine' plants were each unique and identified as 'Turpentine 3', 'Turpentine 7', 'Turpentine 8', 'Turpentine 9', and 'Turpentine 10'. The cultivar Saigon was introduced as seed from

Vietnam. Each seedling has a unique genotype, labeled 'Saigon 1', and 'Saigon 3' through 'Saigon 7'. A unique genotype could not be determined for the cultivar Malindi, represented by two plants with differing genotypes; these were designated as different accessions. A unique cultivar genotype was determined for the remaining 69 cultivars with multiple plants.

Among the 137 single-plant cultivars, seven plants had genotypes that matched with genotypes of multiple plant cultivars, and six were genotypically identical to another six single-plant cultivars resulting in 124 unique genotypes. A total of 208 cultivars were used for further analyses, 203 *M. indica*, three *M. odorata*, and two *M. griffithii* (Table 1). Overall, 22 plants were found to be off types or to be mislabeled. This represents a propagation error of 6.13%.

**Genetic differentiation between horticultural types and relatedness of cultivars.** The UPGMA dendrogram based on Rogers' distance was generated for 10 geographical groups that contain at least four genotypes of *M. indica*, as well as for the two other *Mangifera* species. The dendrogram groups the three *Mangifera* species into separate clusters. The *M. indica* cluster is further divided into four main groups. The populations from Africa, India, and Central America cluster in one group, the Florida, Hawaii, and Israel cultivars in another, the Pacific, South America, and West Indies cultivars in another, with the Southeast Asia cultivars in their own cluster (Fig. 1).

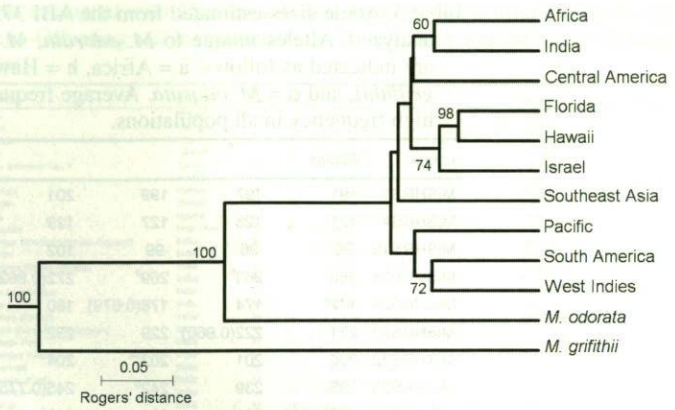


Fig. 1. Unweighted pair group method dendrogram using modified Rogers' distance (Wright, 1978) based on overall microsatellite marker allele frequencies across the 10 main *Mangifera indica* populations by origin and two related *Mangifera* species. Bootstrap values below 50% are not shown.

Comparison by horticultural type (monoembryonic vs. polyembryonic) was accomplished for the 179 cultivars in which this trait has been characterized (Table 1). Of these, 110 are monoembryonic and 69 polyembryonic. Eleven unique alleles were found among the polyembryonic types, all with very low frequency (0.007 to 0.054). Among the monoembryonic types, 34 unique alleles were found, also at very low frequency (0.004 to 0.051), and thus are not useful for the identification of the horticultural type. The average number of alleles was similar between the monoembryonic and polyembryonic types (6.40 and 5.92), as was the $H_{nb}$ (0.571 and 0.596) and the $H_{obs}$ (0.544 and 0.548; Table 4).

Population comparisons by origin indicated that the Florida group is less diverse with a mean number of alleles lower (4.88) than the Indian (5.60) or Southeast Asian groups (5.32) and with all other groups having a lower number of alleles. $H_{nb}$ was higher among the Indian (0.600) and Southeast Asian (0.582) cultivars and lower among the Florida cultivars (0.538). $H_{obs}$ was similar among the Florida and Indian cultivars (0.544 and 0.545, respectively) and slightly lower among the Southeast Asian cultivars (0.538). It was highest in the Pacific cultivars (0.570) and lowest in the Central American cultivars (0.487) (Table 4).

The PCA based on gene frequencies for the Indian, Southeast Asian, and West Indian populations is illustrated in Fig. 2A, along with the PCA of each of these individual populations with the Florida population (Fig. 2B–D). The PCA for the Indian, Southeast Asian, and West Indian populations accounts for 24.8% of the total variation. Two clusters are discernible, one consisting of Southeast Asian cultivars clustering around 'Cambodiana' and the other a mixture of the Indian cultivars, West Indian cultivars, and the Southeast Asian cultivars Tenom, Burma, Kensington, Seinta Lone, and the 'Saigon' genotypes. The 'Saigon' genotypes all cluster closely while the 'Turpentine' land race genotypes have a less cohesive grouping and are distributed throughout the West Indian population (Fig. 2A). The PCA for Florida and Southeast Asia (Fig. 2B) accounts for 28.1% of the total variation and delineates two distinct populations. 'Duncan' clearly groups with the Southeast Asian cultivars. All of the 'Saigon' genotypes along with 'Tenom', 'Burma', and 'Kensington' cluster with the Florida cultivars. The PCA for the Florida and Indian populations (Fig. 2C) accounts for 22.9% of the total variation and reveals considerable overlap of the two groups. The PCA for the Florida and West Indian populations (Fig. 2D) accounts for 25.5% of the total variation and exhibits one major cluster of primarily Florida cultivars and one minor cluster of primarily West Indian cultivars. Some of the West Indian cultivars cluster among the Florida cultivars, these include 'Colombo Kidney' and several of the 'Turpentine' genotypes.

The two accessions of *M. griffithii* contained 12 unique alleles at 11 loci, while the two *M. odorata* accessions contained 15 unique alleles at 12 loci (Table 3). Of the 10 *M. indica* populations categorized by geographic origin only five had unique alleles. Among the African group five unique alleles existed at four different loci, three with a frequency of 0.0625 and two with a frequency of 0.0312. The Indian cultivars contained 13 unique alleles at 11 loci with the highest frequency at mMiCIR014 (allele 165, frequency = 0.0714). The Hawaiian cultivars contained two unique alleles at one locus both with a frequency of 0.0625 while the Pacific cultivars contained a single unique allele at MiSHRS39 (allele 354, frequency = 0.125). Among the Southeast Asian cultivars five unique alleles existed at five different loci of which the highest frequency was at MiSHRS44 (allele 245, frequency = 0.1428). None of the *M. indica* unique alleles had a sufficiently high frequency to be considered population specific (Table 3). Seven loci had alleles at very high frequencies in all populations and these are indicated along with the average frequency across all populations in Table 3.

**Parentage analysis.** Results of the parentage analysis are summarized in Table 5 and a pedigree based on this analysis is illustrated in Fig. 3. The first set included 35 genotypes introduced to Florida between 1880 and 1910. These were the possible contributors to the earliest Florida cultivars of which the two most important are 'Haden' and 'Brooks'. The results of the analysis of Set I indicated that 'Haden' resulted from a 'Mulgoba' x 'Turpentine' hybridization. There were no mismatches between 'Haden' and 'Mulgoba', the known maternal parent. There were two 'Turpentine' genotypes with no mismatches: 'Turpentine 3' and 'Turpentine 10', but Offspring–Candidate parent 1–Candidate parent 2 (O-CP1-CP2) mismatches were revealed for the cross of 'Mulgoba' with either of the 'Turpentine' genotypes. However,

Table 4. Comparison of genetic variation within the two horticultural types (monoembryonic vs. polyembryonic) of *Mangifera indica* and among the main populations by geographic origin across 25 microsatellite loci. Standard deviations are indicated in parentheses for unbiased gene diversity [$H_{nb}$ (Nei, 1978)] and observed heterozygosity ($H_{obs}$).

| Group | Sample size (no. plants) | $P_{0.95}$[z] | Mean alleles/locus (no.) | $H_{nb}$ | $H_{obs}$ |
|---|---|---|---|---|---|
| Type[y] | | | | | |
| Monoembryonic | 110 | 0.920 | 6.400 | 0.571 (0.224) | 0.544 (0.235) |
| Polyembryonic | 69 | 0.960 | 5.920 | 0.596 (0.227) | 0.548 (0.240) |
| Population | | | | | |
| Indian | 29 | 0.960 | 5.600 | 0.600 (0.243) | 0.545 (0.257) |
| Southeast Asian | 34 | 0.960 | 5.320 | 0.582 (0.234) | 0.538 (0.246) |
| Florida | 64 | 0.920 | 4.880 | 0.538 (0.221) | 0.544 (0.244) |
| Hawaii | 9 | 0.920 | 3.920 | 0.595 (0.230) | 0.569 (0.258) |
| Israel | 6 | 0.920 | 3.160 | 0.555 (0.226) | 0.568 (0.322) |
| West Indies | 26 | 0.880 | 4.080 | 0.551 (0.226) | 0.566 (0.272) |
| Central America | 7 | 0.920 | 3.720 | 0.578 (0.257) | 0.487 (0.253) |
| South America | 6 | 0.920 | 3.360 | 0.579 (0.238) | 0.545 (0.279) |
| Pacific | 4 | 0.880 | 2.800 | 0.560 (0.251) | 0.570 (0.350) |
| Africa | 16 | 0.920 | 4.680 | 0.596 (0.222) | 0.543 (0.220) |

[z]Proportion of polymorphic loci when most frequent allele does not exceed 95%.
[y]Of the 203 genotypes studied only 179 were classified as poly- or mono-embryonic.

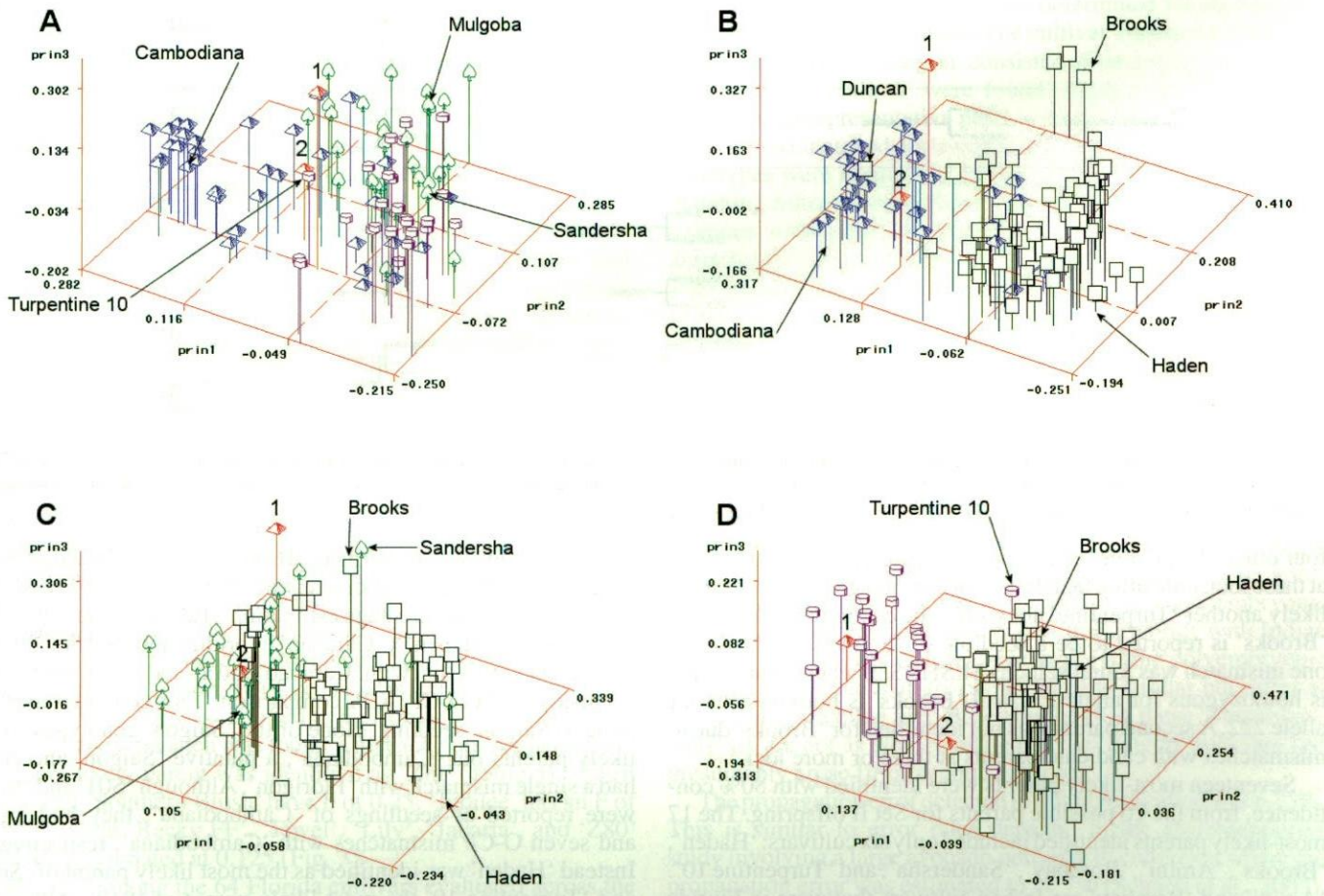


Fig. 2. Principal coordinate analysis (PCA) for the microsatellite evaluation of the mango germplasm. (**A**) PCA for the Indian, Southeast Asian and West Indian cultivars, (**B**) PCA for the Florida and Southeast Asian cultivars, (**C**) PCA for the Florida and Indian cultivars, and (**D**) PCA for the Florida and West Indian cultivars. Cultivar types are represented as follows: Southeast Asian, blue pyramids; Indian, green spades; West Indian, purple cylinders; Florida, black squares; *Mangifera odorata*, red pyramid (1); and *M. griffithii*, red pyramid (2). The cultivars Sandersha, Mulgoba, Turpentine 10, Haden, Brooks, Cambodiana, Duncan, *M. odorata* (1), and *M. griffithii* (2) are included for reference.

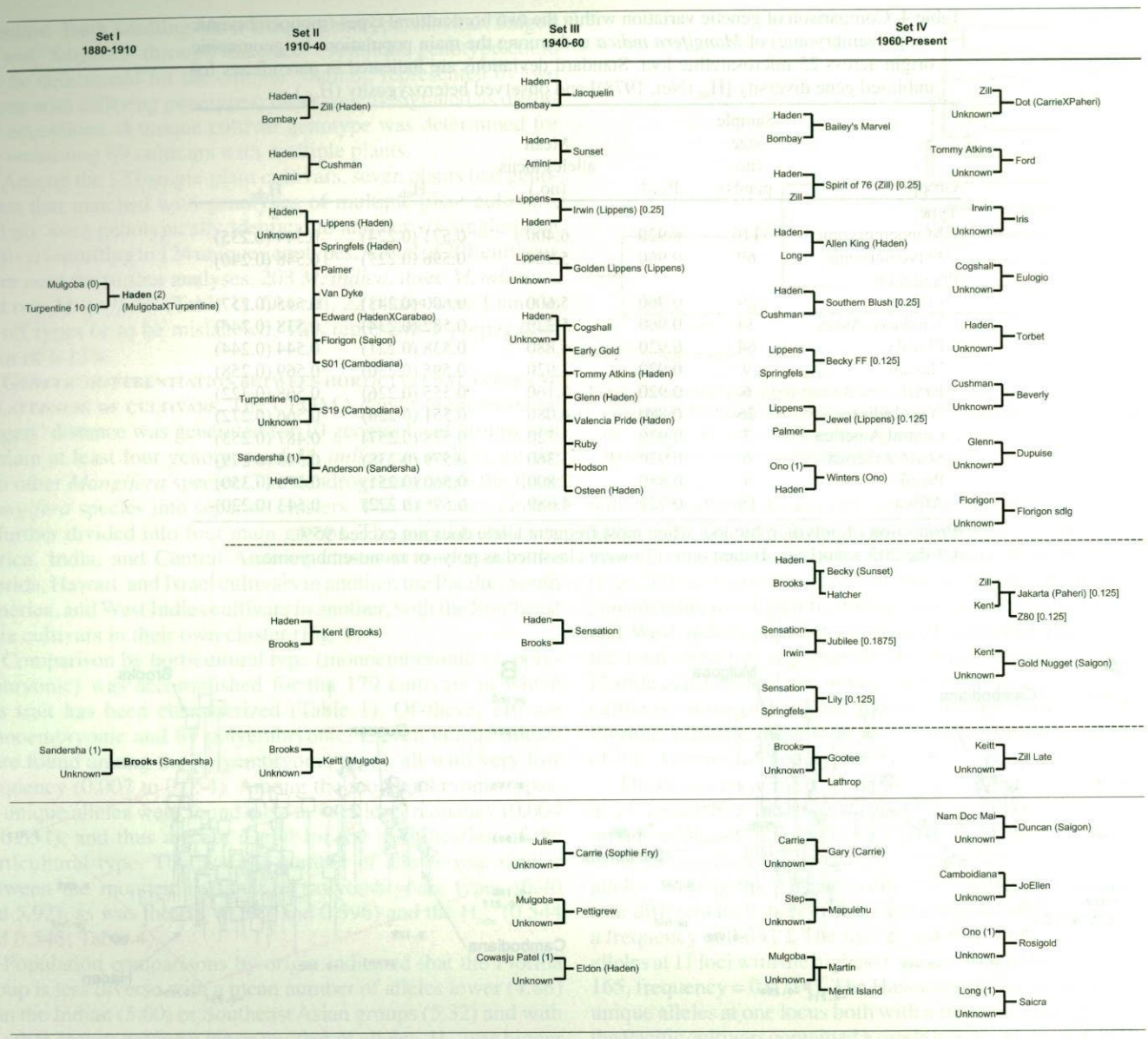


Fig. 3. Pedigree of Florida mango cultivars based on microsatellite allele frequencies estimated using the program CERVUS and a multi-stage parentage selection. The parent reported in the literature is indicated in parenthesis following the Florida cultivar name; the number of loci with offspring-candidate parent mismatches is indicated in parenthesis. The inbreeding coefficient is in brackets.

four other 'Turpentine' genotypes have alleles that would match at these loci, indicating that the actual parental genotype was most likely another 'Turpentine' not within the candidate parent pool. 'Brooks' is reported to be a seedling of 'Sandersha'; however, one mismatch was found at locus MiSHRS30 where 'Sandersha' is homozygous for allele 232 and 'Brooks' is homozygous for allele 222. A second parent was not identified for 'Brooks' due to mismatches with candidate parents at three or more loci.

Seventeen most-likely parents were identified with 80% confidence, from the 26 possible parents for Set II offspring. The 17 most-likely parents identified included only six cultivars: 'Haden', 'Brooks', 'Amini', 'Bombay', 'Sandersha', and 'Turpentine 10'. 'Amini' and 'Bombay' are Indian cultivars that were in Florida after 1900 but did not contribute to 'Haden' or 'Brooks'. All most-likely parents had no O-CP mismatches with the exception of 'Sandersha', which had a single O-CP mismatch with 'Anderson'. Four Florida selections from Set II ('Zill', 'Lippens', 'Springfels', and 'Edward') were expected to have 'Haden' as the maternal parent, and this was confirmed. 'Brooks' was confirmed to be a parent of 'Kent'. Several discrepancies were revealed between historical data and parentage estimation. 'Edward' was said to be the result of a 'Haden' x 'Carabao' hybridization, but there were mismatches at six loci with 'Carabao' and an additional mismatch occurring for the O-CP1-CP2 comparison. 'Florigon' was reported to be a 'Saigon' seedling. None of the 'Saigon' genotypes were likely parents but 'Cambodiana', a putative 'Saigon' ancestor, had a single mismatch with 'Florigon'. Although 'S01' and 'S19' were reported as seedlings of 'Cambodiana', they had eight and seven O-CP mismatches with 'Cambodiana', respectively. Instead 'Haden' was identified as the most likely parent of 'S01' and 'Turpentine 10' the most likely parent of 'S19'. Although a 'Saigon' genotype was not identified as a most-likely parent, allelic configurations indicate that a closely related genotype of the 'Saigon' seedlings could be a parent of 'S01' and 'S19'. The speculation that 'Mulgoba' was the parent of 'Keitt' was unsubstantiated as six O-CP mismatches were found; instead 'Brooks'

was identified as the most-likely parent. With no genotyping errors allowed 'Brooks' was estimated to be the most-likely parent of 'Anderson'. However, after re-analysis with 0.01% genotyping errors allowed the reported parent ('Sandersha') was estimated as the most-likely parent, in spite of one locus mismatch.

Twenty most-likely parents were identified from the 32 possible parents for Set III offspring. The 20 most-likely parents identified consisted of eight cultivars, of which five were common to Set I and Set II. The other three cultivars were: 'Lippens' (a seedling of 'Haden'), 'Julie' (a West Indian cultivar), and 'Cowasju Patel' (an Indian cultivar). The reported parentage of five Florida selections was confirmed. The 'Haden' parentage of 'Eldon' was not confirmed because of two O-CP mismatches. Four candidates had only one O-CP mismatch with 'Eldon', but only one ('Cawasju Patel') with 80% confidence. 'Sophie Fry' is the reported parent of 'Carrie', but this cultivar was not included in the collection and 'Julie' was identified as the most-likely parent of 'Carrie' instead. 'Jacquelin', 'Sensation', 'Sunset', 'Cogshall', 'Hodson', 'Early Gold', and 'Ruby' are of unknown parentage (Campbell, 1992); however, in our analysis 'Haden' was identified as a most-likely parent. In Set III 'Haden' was selected as the most-likely parent 12 times, 'Lippens' twice, 'Brooks' once, with other most-likely parents occuring once ('Mulgoba', 'Bombay', 'Cawasju patel', 'Amini', and 'Julie'). Only 'Eldon' and 'Carrie' did not have 'Haden', 'Brooks', or 'Mulgoba' in their pedigree. The first evidence of inbreeding occurs in Set III where the inbreeding coefficient (F) of 'Irwin' was estimated at 0.25 (Fig. 3).

Set IV included 33 Florida mango cultivars selected between 1960 and 2000. In total 46 of the 66 possible parents were identified, consisting of 23 cultivars. Only seven of the cultivars identified as most-likely parents did not have 'Haden' or 'Brooks' in their pedigrees: 'Long', 'Bombay', 'Nam Doc Mai', 'Carrie', 'Cambodiana', 'Step', and 'Ono'. The remaining 16 cultivars (36 most-likely parents identified) were: 'Haden', 'Brooks', 'Mulgoba', and 13 cultivars derived from 'Haden' and/or 'Brooks'. Reported parentage was confirmed for five Florida selections ('Zill', 'Allen King', 'Jewell', 'Gary', 'Winters'). 'Winters' had one O-CP mismatch with 'Ono', the reported parent and no mismatches with 'Haden'. The reported parentage was not supported for the following selections: 'Becky', 'Duncan', 'Gold Nugget', 'Jakarta', and 'Dot'. 'Becky' was reported to be a seedling of 'Sunset', although mismatches at four loci were detected and 'Becky' was estimated to be from a cross between 'Haden' and 'Brooks'. Among the early mango growers it was common practice to top work seedling trees with scions of both 'Haden' and 'Brooks' and this probably accounts for the misidentification of the maternal parent of 'Becky'. 'Duncan' and 'Gold Nugget' were reported to be 'Saigon' seedlings but all 'Saigon' genotypes were excluded due to mismatches. 'Jakarta' is reported as a seedling of 'Paheri' but five O-CP mismatches were found between them. 'Dot' is reported to be from a 'Carrie' x 'Paheri' cross but one mismatch was found with each putative parent and the more likely parent was identified as 'Zill' (Table 5). 'Spirit of 76' and 'Southern Blush' have F of 0.25, 'Jubilee' has an F of 0.1875, while 'Becky FF', 'Jewel', 'Lily', 'Jakarta', and 'Z80' all have F estimated at 0.125 (Fig. 3).

Overall, among the 64 Florida cultivars evaluated across the four parent-offspring sets, two most-likely parents were identified for 22 cultivars and one most-likely parent for the remaining 42 cultivars; all but one estimation was with 80% confidence. Among the 128 possible parents, 86 (67%) were identified and 42 (33%) remain unknown. 'Haden' was identified as a most likely parent 31 times; 'Brooks' seven times; 'Haden' and/or 'Brooks' derivatives 25 times; 'Mulgoba', the maternal parent of 'Haden', four times and 'Sandersha', the maternal parent of 'Brooks', two times. The remaining 25 most-likely parents were six Indian (three cultivars), five West Indian (three cultivars), two Southeast Asian (two cultivars), and one Florida selection of West Indian descent.

## Discussion

The microsatellite loci in this study were moderately polymorphic with an average of 6.96 alleles per locus and average PIC value of 0.552. Both of these values are higher than those reported by Viruel et al. (2005; 5.3 and 0.28), but they evaluated 28 genotypes vs. the 203 unique *M. indica* genotypes evaluated in this study. The level of polymorphism detected among the mango accessions is considerably lower than estimates from a similar collection of avocado (*Persea americana* Mill.) in which the average number of alleles was 18.8 and $H_{nb}$ was 0.83 (Schnell et al., 2003). Both of these species are assumed to be highly outcrossed and highly heterozygous. Avocado does have a distinctive reproductive system of protogynous, diurnally synchronous dichogamy (Bergh, 1969), which promotes cross pollination, and this may explain the differences between results from these two tropical fruit species.

A unique cultivar genotype was determined for 69 of the 72 cultivars with multiple plants. The cultivar Turpentine consisted of seven genotypes, 'Saigon' consisted of six genotypes, and two genotypes of 'Malindi' were found. Among the 137 cultivars with a single representative plant, seven matched multiple plant cultivars, six matched other single plant cultivars, and 124 unique genotypes were found. We chose a single name for each of the matching cultivars and both are listed in Table 1. The single-plant cultivars with genotypes matching that of other cultivars were either found to represent alternative names for the same cultivar, or were considered propagation errors. Four cultivars with multiple plants were problematic for accession genotype determination. Initially, a unique genotype could not be determined for the cultivars Florigon or Malindi; each is represented by two plants with differing genotypes. Upon searching the database records one 'Florigon' genotype was found to have been obtained as seed from the original 'Florigon' cultivar. This was confirmed by the parentage analysis: one of the 'Florigon' accessions was identified as the most-likely parent of the other. The two 'Malindi' were labeled 'Malindi 1' and 'Malindi 2'. The cultivar Saigon was introduced as seed from Vietnam and each of the six seedlings has a unique genotype. This cultivar is monoembryonic and the PCA suggests that its parents were hybrids with Indian types. We have seven different genotypes of the polyembryonic cultivar Turpentine that are phenotypically similar but differ at several microsatellite loci. Based on the PCA these 'Turpentine' cultivars are not as closely related as the 'Saigon' cultivars and presumably arose from zygotic embryos.

The propagation error detected in this investigation was 6.13%. This is similar to error rates detected in the aforementioned study involving a large avocado germplasm collection where the propagation error was estimated to be 7% (Schnell et al., 2003). Both of these are far lower than propagation errors detected in international cacao (*Theobroma cacao* L.) germplasm collections, estimated to be >30% (Motilal et al., 2002). These errors could be classified into two types. The first type consisted of plants with the same cultivar name but with a different microsatellite

Table 5. Results of the parentage analysis of Florida *Mangifera indica* cultivars performed using the program CERVUS, based on 25 SSR loci. In each of the four sets, divided chronologically, the most likely parent (Candidate parent 1) is identified for each Florida selection (offspring). A second most likely parent (Candidate parent 2) was identified for some of the Florida selections.

| Florida selection (offspring) | Candidate parent 1 | Offspring-Candidate parent 1 loci mismatches (no.) | Candidate parent 2 | Offspring-Candidate parent 2 loci mismatches (no.) | Offspring-Candidate parent 1 - Candidate parent 2 loci mismatches (no.) | LOD[z] | Delta[y] | Confidence[x] |
|---|---|---|---|---|---|---|---|---|
| **Set I** | | | | | | | | |
| Haden | Mulgoba | 0 | Turpentine 10 | 0 | 2 | 1.01E+01 | 6.13E+00 | * |
| Brooks | Sandersha | 1 | - | - | - | 5.86E+00 | 5.77E+00 | * |
| **Set II** | | | | | | | | |
| Cushman | Haden | 0 | Amini | 0 | 0 | 7.49E+00 | 7.49E+00 | * |
| Kent | Brooks | 0 | Haden | 0 | 0 | 1.26E+01 | 1.26E+01 | * |
| Zill | Bombay | 0 | Haden | 0 | 0 | 1.26E+01 | 1.26E+01 | * |
| Anderson | Sandersha | 1 | Haden | 0 | 1 | 7.36E+00 | 3.56E+00 | * |
| Edward | Haden | 0 | - | - | - | 4.20E+00 | 4.20E+00 | * |
| Keitt | Brooks | 0 | - | - | - | 7.40E+00 | 7.40E+00 | * |
| Lippens | Haden | 0 | - | - | - | 5.50E+00 | 5.50E+00 | * |
| Florigon | Haden | 0 | - | - | - | 2.62E+00 | 2.62E+00 | * |
| Palmer | Haden | 0 | - | - | - | 2.38E+00 | 2.30E+00 | * |
| S01 | Haden | 0 | - | - | - | 4.19E+00 | 4.19E+00 | * |
| S19 | Turpentine 10 | 0 | - | - | - | 7.83E+00 | 3.19E+00 | * |
| Springfels | Haden | 0 | - | - | - | 6.35E+00 | 4.06E+00 | * |
| Van Dyke | Haden | 0 | - | - | - | 3.35E+00 | 1.26E+00 | * |
| **Set III** | | | | | | | | |
| Irwin | Haden | 0 | Lippens | 0 | 0 | 6.55E+00 | 3.08E+00 | * |
| Jacquelin | Bombay | 0 | Haden | 0 | 0 | 8.01E+00 | 8.01E+00 | * |
| Sensation | Brooks | 0 | Haden | 0 | 0 | 1.08E+01 | 1.08E+01 | * |
| Sunset | Amini | 0 | Haden | 0 | 0 | 9.74E+00 | 9.74E+00 | * |
| Carrie | Julie | 0 | - | - | - | 1.13E+01 | 1.13E+01 | * |
| Golden Lippens | Lippens | 0 | - | - | - | 8.10E+00 | 8.10E+00 | * |
| Pettigrew | Mulgoba | 0 | - | - | - | 7.81E+00 | 7.81E+00 | * |
| Tommy Atkins | Haden | 0 | - | - | - | 5.27E+00 | 2.00E-01 | * |
| Cogshall | Haden | 0 | - | - | - | 3.90E+00 | 6.88E-03 | * |
| Earlygold | Haden | 0 | - | - | - | 1.94E+00 | 1.65E+00 | * |
| Glenn | Haden | 0 | - | - | - | 4.24E+00 | 2.48E+00 | * |
| Hodson | Haden | 0 | - | - | - | 2.39E+00 | 5.80E-01 | * |
| Ruby | Haden | 0 | - | - | - | 4.13E+00 | 2.58E+00 | * |
| Valencia Pride | Haden | 0 | - | - | - | 1.83E+00 | 1.29E+00 | * |
| Eldon | Cowasju Patel | 1 | - | - | - | 8.04E+00 | 1.89E+00 | * |
| Osteen | Haden | 0 | - | - | - | 3.25E+00 | 0.00E+00 | - |
| **Set IV** | | | | | | | | |
| Allen-King / Everbearing | Haden | 0 | Long | 0 | 0 | 1.16E+01 | 3.98E+00 | * |
| Bailey's Marvel | Bombay | 0 | Haden | 0 | 0 | 6.57E+00 | 3.07E+00 | * |
| Becky | Brooks | 0 | Haden | 0 | 0 | 5.97E+00 | 3.93E+00 | * |
| Becky FF | Lippens | 0 | Springfels | 0 | 0 | 9.36E+00 | 8.38E+00 | * |
| Hatcher | Brooks | 0 | Haden | 0 | 0 | 1.02E+01 | 1.53E+00 | * |
| Jakarta | Kent | 0 | Zill | 0 | 0 | 9.27E+00 | 6.18E+00 | * |
| Jewel | Lippens | 0 | Palmer | 0 | 0 | 9.87E+00 | 9.87E+00 | * |
| Jubilee | Sensation | 0 | Irwin | 0 | 0 | 7.96E+00 | 3.03E+00 | * |
| Lily | Springfels | 0 | Sensation | 0 | 0 | 9.21E+00 | 7.72E+00 | * |
| Southern Blush | Haden | 0 | Cushman | 0 | 0 | 7.67E+00 | 1.00E-01 | * |
| Spirit of 76 | Zill | 0 | Haden | 0 | 0 | 7.82E+00 | 1.84E+00 | * |
| Z80 | Kent | 0 | Zill | 0 | 0 | 9.76E+00 | 4.61E+00 | * |
| Winters | Ono | 1 | Haden | 0 | 1 | 8.87E+00 | 2.69E+00 | * |
| Duncan | Nam Doc Mai | 0 | - | - | - | 1.49E+01 | 3.63E+00 | * |
| Ford | Tommy Atkins | 0 | - | - | - | 5.30E+00 | 1.60E+00 | * |
| Gary | Carrie | 0 | - | - | - | 8.60E+00 | 5.89E+00 | * |
| Golden Nugget | Kent | 0 | - | - | - | 5.57E+00 | 1.08E+00 | * |
| Gootee | Brooks | 0 | - | - | - | 9.80E+00 | 5.19E+00 | * |
| Iris | Irwin | 0 | - | - | - | 4.76E+00 | 2.71E+00 | * |
| Joellen | Cambodiana | 0 | - | - | - | 6.73E+00 | 6.57E+00 | * |
| Lathrop | Brooks | 0 | - | - | - | 1.05E+01 | 5.29E+00 | * |
| Mapulehu | Step | 0 | - | - | - | 9.59E+00 | 7.07E-01 | * |
| Rosigold | Ono | 1 | - | - | - | 9.34E+00 | 7.71E+00 | * |
| Saiacra | Long | 1 | - | - | - | 1.05E+01 | 6.07E+00 | * |
| Torbet | Haden | 0 | - | - | - | 5.10E+00 | 6.07E-01 | * |
| Zilate | Keit | 0 | - | - | - | 5.80E+00 | 1.87E+00 | * |
| Beverly | Cushman | 0 | - | - | - | 3.27E+00 | 5.28E-01 | * |
| Dot | Zill | 0 | - | - | - | 7.71E+00 | 7.71E+00 | * |
| Eulogio | Cogshall | 0 | - | - | - | 5.41E+00 | 1.45E+00 | * |
| Martin | Mulgoba | 0 | - | - | - | 8.25E+00 | 8.25E+00 | * |
| Merrit Island | Mulgoba | 0 | - | - | - | 8.63E+00 | 8.63E+00 | * |
| Florigon sdlg | Florigon | 0 | - | - | - | 2.93E+00 | 2.93E+00 | * |
| Dupuise | Glenn | 1 | - | - | - | 3.45E+00 | 6.04E-01 | * |

[z]LOD = sum of the log-likelihood ratios at each locus.
[y]Difference in LOD scores between the most likely candidate parent and the second most likely candidate parent.
[x]* = 80 %; - = most likely (<80%).

genotype. The second type consisted of plants with a different cultivar name but with identical microsatellite genotypes. Ten of the errors were of the first type and nine of these were very clear with mismatches at two or more loci. The other one had a mismatch at one locus that could have been the result of genotyping error. For the second type, 12 errors were detected and all could be confirmed as mislabeling based on phenotypic traits.

Comparisons between polyembryonic and monoembryonic types did not reveal significant differences between gene frequencies. No unique alleles were found that could be used for classification of types. All of the Indian cultivars classified (24) were monoembryonic. Of the 33 Southeast Asian cultivars classified, 25 were polyembryonic; 'Ponsawade', 'Mammou', and the six 'Saigon' genotypes, were monoembryonic. Fifty-eight of the Florida cultivars have been classified with 50 being monoembryonic and eight polyembryonic (Table 1). Aron et al.

(1998) demonstrated polyembryony was dominant in mango. 'Haden' is a cross of the monoembryonic cultivar Mulgoba and the polyembryonic cultivar Turpentine. If we assume that a single dominant gene controls this trait, all of the Indian cultivars must be homozygous recessive and the 'Turpentine' parent of 'Haden' must have been heterozygous. The evidence suggests that 'Haden' inherited the recessive allele from 'Turpentine', as all identified progeny of 'Haden' are monoembryonic with the exception of 'Winters'. The most-likely second parent of 'Winters' is 'Ono', a polyembryonic cultivar from Hawaii. The frequency of this dominant allele is low in the Florida population and absent from the Indian cultivars. The early introductions of polyembryonic cultivars were of inferior quality, but they did flower and fruit under south Florida conditions. In contrast, the Indian cultivars imported by the USDA did not flower or set fruit dependably nor did they contain adequate disease resistance. Hybridization of the Indian cultivars with the polyembryonic types was followed by selection for the Indian fruit characteristics in seedlings that would flower and set fruit dependably. This has resulted in most of the Florida cultivars being monoembryonic.

Genetic distances clearly separate *M. indica* populations into four clusters and separate the Southeast Asian cultivars from all other cultivars (Fig. 1). The close relationship between the Florida, Israel, and Hawaii cultivars is also indicated and this is expected as much germplasm was exchanged between these three areas. The relationship of the Florida cultivars with the Indian, Southeast Asian, and West Indian cultivars is also illustrated in Fig. 2 where the Southeast Asian cultivars are clearly delineated from the other groups in Fig. 2A and 2B. Both analyses support a closer association between the Indian and Florida cultivars when compared to the Southeast Asian cultivars.

Pedigree reconstruction depends on availability, accuracy, and completeness of historical data as well as accuracy and robustness of genetic data. The results of the parentage analysis yield the best approximation for the pedigree, given the data available. Discrepancies between parentage reported in the literature and that based on molecular marker estimations may arise due to errors in record keeping, incomplete data (not all true parents were sampled) or errors in propagation, as well as in genotyping due to null alleles or mutations. For the Florida cultivars in this study, 20 of the reported parents in the literature were confirmed. 'Brooks' is a cultivar selected in Florida, reported to be a seedling of 'Sandersha' planted on the property of Mr. Brooks in Miami. All candidate parents were excluded for 'Brooks' when no genotyping error was allowed. However, with 0.01% error allowed 'Sandersha' was identified as the most-likely parent with one O-CP mismatch at locus MiSHRS30. This could result from the presence of a null allele in 'Sandersha' that was inherited by 'Brooks'. Both 'Sandersha' and 'Brooks' are homozygous for different alleles at MiSHRS30. In addition, 'Anderson' was said to have been grown from seed of 'Sandersha' sent from Jamaica and 'Sandersha' is recorded as having been sent to Jamaica by David Fairchild. 'Sandersha' was identified as the most-likely parent with one O-CP mismatch at locus MiSHRS30, where both are homozygous for different alleles. However, when no genotyping errors are allowed, 'Brooks' is identified as the most-likely parent of 'Anderson' with no mismatches. It is a more likely scenario that seed from 'Sandersha' rather than 'Brooks' would have been sent from Jamaica at that time period (FMF, 1948). The null allele hypothesis can also apply to the mismatch at locus MiSHRS32 existing between 'Winters' (homozygous for allele 204) and its reported parent 'Ono' (homozygous for allele 201). Several other discrepancies between reported parents and estimated parents involved greater numbers of mismatches that cannot be explained by null alleles. 'Mulgoba' was not identified as a likely parent for 'Keitt' due to six O-CP mismatches. Instead, 'Brooks' was identified as the most-likely parent, as it was found to have no mismatches. For the Florida selection 'Carrie', 'Julie' was identified as the most-likely parent in the absence of 'Sophie Fry', the reported parent. 'Julie' is the reported parent of 'Sophie Fry' and thus would actually be a grandparent to 'Carrie'. 'Dot' is reported as resulting from a 'Carrie' x 'Paheri' cross, but with one mismatch for each, was estimated instead to have 'Zill' as a most-likely parent. In this study 'Paheri' and 'Bombay' were synonymous and 'Bombay' was found to be a most-likely parent of 'Zill'. The 'Haden' parentage of 'Glenn' and 'Osteen' was confirmed in this study, in contrast to the RAPD analysis where this parentage was questioned (Schnell et al., 1995). However, the RAPD study suggested 'Haden' parentage for 'Eldon' that is not confirmed here. This demonstrates the advantage of using microsatellites, which are genetically defined loci vs. the random amplification of undefined DNA fragments.

Among the 64 Florida cultivars evaluated in the parentage analysis the genetic background was found to be based on as few as four Indian cultivars, and the polyembryonic cultivar Turpentine. Two Indian cultivars, 'Mulgoba' and 'Sandersha', are in the background of most Florida types with 'Amini', 'Bombay', 'Cambodiana', 'Long', 'Julie', and 'Nam doc mai' making lesser contributions. In the parentage analysis 'Turpentine 10' was identified as a most probable paternal parent for 'Haden'; however, two O-CP1-CP2 mismatches were detected. The paternal parent for 'Haden' is most likely another 'Turpentine' genotype not in our collection. The seedling races of Cuba and Florida were considered the same by Popenoe (1920) who called them the West Indian race commonly known as 'Turpentine' in Florida. Based on the PCA, they represent a diverse group of genotypes (Fig. 2). 'Haden' is reported as the maternal parent for 10 cultivars included in this analysis, but based on the parentage analysis we found 31 cultivars with 'Haden' as one of the most-likely parents. Likewise, the other important early Florida selection 'Brooks' is the parent of seven cultivars. 'Haden, 'Brooks', and seedlings of 'Haden' and 'Brooks' have contributed disproportionately to the Florida group. Many minor Florida cultivars, known from the literature but not in the germplasm collections, were not included in this study and are likely derivatives of 'Haden' and 'Brooks' (Campbell, 1992; Ruehle and Ledin, 1955). If these are identified as the second parent for Florida cultivars in Set III and IV the degree of inbreeding may be higher than estimated.

Florida has been considered a secondary center of diversity for mango as many Indian and Southeast Asian cultivars were imported and the Florida cultivars were developed from these imported cultivars (Knight and Schnell, 1994; Mukherjee, 1997; Schnell et al., 1995). A substantial amount of genetic diversity exists in germplasm collections in southern Florida; however, the Florida cultivars are not more diverse than the Indian or Southeast Asian cultivars. Florida may not truly represent a secondary center for diversity of the species as has been previously reported. Our data (Fig. 2 C and D) suggest that the Florida cultivars represent hybrids between a few Indian and West Indian cultivars selected for fruit phenotypes more commonly found among the Indian cultivars, as well as for productivity under southern Florida conditions.

Selecting under southern Florida conditions has led to a group of cultivars including 'Keitt', 'Tommy Atkins', 'Haden', 'Parvin', and 'Irwin' that produce dependably over a range of environmen-

tal conditions. This stability found among the Florida cultivars merits further investigation as it also occurs in other fruit species. Understanding this valuable genetic architecture, how it arose, and its mode of inheritance will be important for future mango breeding efforts when Florida cultivars are used as parents for the selection of new cultivars.

### Literature Cited


Adato, A., D. Sharon, and U. Lavi. 1995. Application of DNA fingerprints for identification and genetic analyses of mango (*Mangifera indica*) genotypes. J. Amer. Soc. Hort. Sci. 120:259–264.

Aron, Y., H. Czosnek, and S. Gazit. 1998. Polyembryony in mango (*Mangifera indica* L.) is controlled by a single dominant gene. HortScience 33:1241–1242.

Bergh, B.O. 1969. Avocado (*Persea americana* Mill.), p. 23–51. In: F.P. Ferwerda and F. Wit (eds.). Outlines of perennial crop breeding in the tropics. Misc. Paper 4 Landbouwhogeschool, Wageningen, The Netherlands.

Cambell, R.J. 1992. A guide to mangoes in Florida. Fairchild Tropical Garden, Miami, Fla.

Degani, C., R. El-Batsri, and S. Gazit. 1990. Enzyme polymorphism in mango. J. Amer. Soc. Hort. Sci. 115:844–847.

Duval, M.F. J. Bunel, C. Sitbon, and A.M. Risterucci. 2005. Development of microsatellite markers for mango (*Mangifera indica* L.). Mol. Ecol. Notes.

Eiadthong, W., K. Yonemori, A. Sugiura, N. Utsunomiya, and S. Subhadrabandhu. 1999. Identification of mango cultivars of Thailand and evalulation of their genetic variation using the amplified fragments by simple sequence repeat-(SSR-) anchored primers. Scientia Hort. 82:57–66.

Florida Mango Forum. 1948. Proceedings of the Florida Mango Forum II. 30 July 1948. Forida Mango Forum, Stuart, Fla.

Florida Mango Forum. 1951. Mango studies. Florida Mango Forum, Stuart, Fla.

Gower, J.C. 1966. Some distance properties of latent root and vector methods used in multivariate analysis. Biometrika 53:325–338.

Kashkush, K., F. Jinggui, E. Tomer, J. Hillel, and U. Lavi. 2001. Cultivar identification and genetic map of mango (*Mangifera indica*). Euphytica 122:129–136.

Knight, R.J. and R.J. Schnell. 1994. Mango introduction in Florida and the 'Haden' cultivar's significance to the modern industry. Econ. Bot. 48:139–145.

Langella, O. 2002. Populations v. 1.2.28, Population genetic software. 22 June 2005 <http://www.pge.cnrs-gif.fr/bioinfo/populations/index.php?lang=en&doc=populations>.

Li, Y.C., A.B. Korol, T. Fahima, A. Beiles, and E. Nevo. 2002. Microsatellites: Genomic distribution, putative functions and mutational mechanisms: A review. Mol. Ecol. 11:2453–2465.

Liu, B-H. 1998. Stastistical genomics. Linkage, mapping and QTL analysis. CRC Press, Boca Raton Fla.

Lopez-Valenzuela, J.A., O. Martinez, and O. Paredes-Lopez. 1997. Geographic differentiation and embryo type identification in *Mangifera indica* L. cultivars using RAPD markers. HortScience 32:1105–1108.

Marshall, T.C., J. Slate, L. Kruuk, and J.M. Pemberton. 1998. Statistical confidence for likelihood-based paternity inference in natural populatons. Mol. Ecol. 7:639–655.

Motilal, L.A., D.R. Butler, and V. Mooledhar. 2002. Verification in global cacao germplasm collections. Ingenic Nwsl. 7:4–8.

Mukherjee, S.K. 1997. Introduction: Botany and importance, p. 1–19. In: R.E. Litz (ed.). The mango: Botany, production and uses. CAB Intl., New York.

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

Nei, M. 1987. Molecular evolutionary genetics. Colombia Univ. Press, New York.

Popenoe, W. 1920. Manual of tropical and subtropical fruits. Macmillan, New York.

Ruehle, G.D. and R.B. Ledin. 1955. Mango growing in Florida. Univ. Fla. Agr. Expt. Sta. Bul. 574.

Schnell, R.J., J.S. Brown, C.T. Olano, E.J. Power, C.A. Krol, D.N. Kuhn, and J.C. Motamayor. 2003. Evaluation of avocado germplasm using microsatellite markers. J. Amer. Soc. Hort. Sci. 128:881–889.

Schnell, R.J., R. Goenaga, and C.T. Olano. 1999. Genetic similarities among cocoyam cultivars based on randomly amplified polymorphic DNA (RAPD) analysis. Scientia Hort. 80:267–276.

Schnell, R.J., C.T. Olano, W.E. Quintanilla, and A.W. Meerow. 2005. Isolation and characterization of 15 microsatellite loci from mango (*Mangifera indica* L.) and cross-species amplification in closely related taxa. Mol. Ecol. Notes 5:625–627.

Schnell, R.J., C.M. Ronning, and R.J. Knight, 1995. Identification of cultivars and validation of genetic relationships in *Mangifera indica* L. using RAPD markers. Theor. Appl. Genet. 90:269–271.

Singh, L.B. 1960. The mango. Interscience, New York.

Slate, J., T. Marshall, and J.M. Pemberton. 2000. A retrospective assessment of the accuracy of the paternity inference program Cervus. Mol. Ecol. 9:801–808.

Viruel, M.A., P. Escribano, M. Barbieri, M. Ferri, and J.I. Hormaza. 2005. Fingerprinting, embryo type and geographic differentiation in mango (*Mangifera indica* L., Anacardiaceae) with microsatellites. Mol. Breeding 15:383–393.

Wright, S., 1978. Evolution and the genetics of populations, Vol. 4. Variability within and among natural populations. Univ. of Chicago Press, Chicago.