Use of DNA Markers for Fingerprinting Compact Clones and Determining the Genetic Relationship between *Elaeis oleifera* Germplasm Origins

E. Araya[1], A. Alvarado[1], R. Escobar[1]

ABSTRACT

ASD's oil palm breeding program started in the 1960s after several genetically advanced and open pollination (wild germplasm) populations were introduced to Costa Rica. However, the use of molecular biology was recently implemented to provide the well-established breeding program with molecular tools such as DNA markers. The objective is to improve the selection process of superior parental palms in Elaeis guineensis *and* Elaeis oleifera *germplasm, seeking the development of new more productive oil palm varieties.*

ASD's molecular biology program started in 2006 through an agreement with the Technological Institute of Costa Rica (ITCR). Simultaneously, ASD established a brand new molecular biology laboratory in Coto, Costa Rica in 2009. The work done at the biotechnology facilities in the ITCR enabled the adaptation of current and routinely used protocols for DNA isolation, genomic microsatellites and AFLPs analysis in the new lab; fragment analyses of molecular markers are now performed by capillary electrophoresis in the ABI 3130 DNA sequencer (Applied Biosystems). ASD has been developing several applications in the breeding program through the use of these techniques, but only the fingerprinting of ASD commercial compact clones and the study of genetic relationships in ASD's Elaeis oleifera germplasm are discussed in this paper.

A reliable and reproducible fingerprinting was developed, by amplifying the DNA of 20 clones in two different laboratories (ITCR and ASD) with the same 17 genomic microsatellites previously described to detect SSRs in E. guineensis *and* E. oleifera*. Microsatellite fragment analysis by capillary electrophoresis was performed in two different ABI 3130 devices belonging to the Molecular and Cellular Biology Research Center of the University of Costa Rica and to ASD. The use of identical methodologies in this side by side experiment yielded the same genetic profile of clones with a multiloci probability of identity of 3.60 x 10^{-13} and a probability of exclusion >0.99999. We obtained 4.05 ± 1.39 alleles per locus with an average expected heterozygosity (H_e) of 0.55, plus three microsatellites amplified double loci which were not previously reported. Principal Coordinates Analysis (PCoA) on the first two axes accounted for 46.3% of the variation observed and the genetic relationship between clones clearly showed the grouping of clones according to their pedigree. These results revealed that genomic microsatellites are suitable for evaluating the legitimacy of clones; future applications include the quality control of clone production to guarantee genetic uniformity and the*

[1] ASD Costa Rica (Agricultural Services & Development), PO Box 30-1000, San Jose, Costa Rica

use of fingerprinting for germplasm registration and genetic relatedness between different germplasm origins.

The large ASD collection of Elaeis oleifera *germplasm was analyzed using 8 genomic microsatellites.* Oleifera *from Honduras, Nicaragua, Costa Rica, Panama, Colombia, Surinam, Ecuador and Brazil were evaluated. PCoA accounted for 80.61% of molecular variation and showed a clear separation of the genetic material according to geographical origin as previously reported. The germplasm from Brazil (Manaus origin) and the progenies derived from this source exhibited the highest diversity with a* H_e *>0.47, related to* oleifera *from Ecuador. On the other hand, Central American/Colombian germplasm had a low genetic differentiation and PCoA characterized these germplasms as a close-related group without distinction between Central America, Panama and Colombia origins, for both* F_1 *and* F_2 *generations. The lowest diversity was detected in Surinam (*H_e*=0.16), however exclusive allele amplification allowed a large differentiation of this germplasm from the other origins.*

Also AFLP analyses with three primer combinations of enzymes EcoRI/MseI in the same germplasm were performed. The results with AFLPs suggested a low genetic diversity with a mean heterozygosis ranging from 0.08 to 0.20. However, polymorphic fragments scored were used in PCoA, where the first two axes represented 72.78% of the total molecular variation. Consistently with SSRs, Brazilian and Surinam oleifera *remained separated and Ecuador was clearly distinct from those of Brazil origin. Furthermore, AFLPs allowed the distinction of the Central American/Colombian populations into three groups:* F_2 *breeding populations with genetic elements from Costa Rica, Panama and Colombia;* F_1 *populations with Central American germplasm involved; the third group was composed of the Colombian* F_1 *population and Costa Rica germplasm crossed with Colombian. The analyses of* E. oleifera *with both DNA markers revealed the highest divergence of Brazilian origin with the combined germplasm from Central America, Colombia, Ecuador and Surinam. Hence, the widening of genetic base and heterozygosis increase might be reached by crossing Central American/Colombian* oleifera *with Brazilian ones, considering only the relevant phenotypic traits from each origin. On the other side, inter-crossing Central American populations may result in more homozygous populations with allele fixation.*

INTRODUCTION

ASD Costa Rica is an oil palm research and technology company established in Costa Rica, Central America that has been developing commercial oil palm varieties characterized by high fresh fruit bunch and high oil yield since 1970. The research program at ASD focuses on three main aspects: the use of plant breeding to create new varieties, the development of tissue culture technologies and agricultural practices research.

The oil palm breeding program at ASD focuses on i) high yielding conventional varieties; ii) non-conventional varieties, including those adapted to marginal growing conditions such as drought and low temperatures, tolerant disease varieties and with high

oleic acid content; and iii) compact varieties and clones for high density planting and longer commercial life span. Cloning technology might be considered the last phase in a selection process of superior palms derived from the conventional breeding program.

The genetic base of ASD's *Elaeis oleifera* is related to a large germplasm collection established during the 1960s and 70s. Many accessions were made during this period by prospecting in several wild populations in regions of Honduras, Nicaragua, Costa Rica, Panama, Colombia, Surinam and Brazil (Escobar 1981). More than 350 accessions were prospected in almost 40 locations and several of them were evaluated in OxG progeny tests in the 1978 plantations (Alvarado *et al*. 2009; Sterling *et al*. 1999). In 2003, this collection was increased with four germplasm sources introduced from Taisha, Ecuador. Most of the *E. oleifera* germplasm was planted in Coto, Costa Rica, and part of this collection was exchanged for *Elaeis guineensis* germplasm from research stations in Africa and Asia during the 1970s.

The creation of new oil palm varieties using conventional breeding methodologies required major experiment evaluation efforts over several years, in different locations (GxE interactions) and extensive areas (Rajanaidu *et al.* 1993, Billotte *et al*. 2001). The limitations of traditional breeding methods might be overcome using molecular biology techniques that allow early selection of desired genotypes, knowledge of the genetic structure of populations, identification and determination of heterozygosity levels in selected progenies and the relationship of desired traits to their genotypes (Billotte *et al.* 2001). Molecular biology tools such as DNA markers and functional genomics are used to study the genotypes and their interaction with phenotypes.

Molecular markers research in oil palm (*Elaeis guineensis*) has been used to study the genetic variation in germplasm collections using RAPDs, microsatellites and isozymes (Shah *et al.* 1994; Shah and Nyuk 1996; Hayati *et al.* 2004); allelic diversity was also determined in natural *E. guineensis* populations in Africa (Bakoume *et al.* 2006). In oil palm breeding programs the genetic structure of germplasm was studied with RFLPs (Mayes *et al.* 2000), and genetic distance and variation were investigated in progenies and their parents through the use of SSR, AFLP and isozymes (Norziha *et al.* 2007; Purba *et al*. 2000).

The genetic diversity in *E. oleifera* was also studied with molecular markers. AFLPs and RFLPs analyses confirmed the identification of the four principal genetic groups in the American oil palm species: Brazil, French Guyana/Surinam, Peru and Central America; these groups were widely separate from *E. guineensis* (Barcelos *et al.* 2002; Billotte *et al.* 2001). The molecular characterization of *E. oleifera* with SSRs from the Amazonian areas in Colombia was complemented with agronomical and biochemical analyses (Rocha y Rey 2007) and RAPD markers revealed high genetic variability within the populations or accessions collected in the Amazonian forest (Moretzsohn *et al.* 2002). On the other hand, the development of genetic profiles (fingerprinting) of *E. guineensis* was used to confirm the genetic legitimacy of clones using RFLP markers (Jack *et al*. 1998). The set of microsatellites developed by Singh *et al.* (2007) proved useful for identifying clones, clonal mixtures detection, monitoring the uniformity of lines *in vitro* and confirming the identity of ramets for recloning.

ASD has recently ventured into the molecular biology area by establishing a laboratory where protocols have been implemented to use microsatellites and AFLP markers. The scientific activities of ASD's molecular biology program were initiated in 2006, thanks to an agreement with the Technological Institute of Costa Rica (ITCR). The ASD molecular biology laboratory in Coto, Southern Costa Rica, near the ASD experimental station began activities in 2009. The objective of this laboratory is to provide the conventional breeding program with molecular tools that will help achieve more efficient selection of superior parental palms or lines.

Several applications of the molecular markers are being implemented; however, in this document only two applications are presented: the molecular identification of oil palm for fingerprinting and parentage analysis of clones and the study of genetic relationships within ASD's *Elaeis oleifera* germplasm. In addition, some results of the optimization process are presented in order to demonstrate the reliability of the results.

MATERIALS AND METHODS

DNA isolation

The type, age and source of sampling tissue are important to consider when performing DNA isolation, mainly because the younger leaf tissues possess higher photosynthetic activity and continued growth rates, which favor the accumulation of polysaccharides. Therefore, DNA extraction was performed in young palms with CTAB and phenol by modifying the methodology described by Rocha (2002); instead, mini-DNA preparations with reduced volumes were performed. In older palms, DNA was extracted with CTAB using a methodology that has shown to be effective for obtaining high quality DNA in several species including *E. guineensis* (Araya *et al.* 2005a; Araya *et al.* 2005b; Aguero, 2007, Fernández 2008).

Microsatellites and AFLPs analyses

The preliminary work with microsatellites was done with 26 primers previously developed by CIRAD (Table 1); 21 were developed by Billotte *et al.* (2001) and amplify SSRs in *Elaeis guineensis* and *Elaeis oleifera*; the other five were described by Billotte *et al.* (2005) in *Elaeis guineensis*. The detection of good quality PCR products and the reduction of nonspecific amplifications were reached by optimization of annealing temperatures (Tm) on primers. These allowed acquisition of very clear and defined electropherograms and banding patterns for genotyping of samples on the ABI 3130 genetic analyzer (Applied Biosystems) and acrylamide gels, respectively.

So far, the main results in the molecular biology program have been obtained with microsatellites; nevertheless, optimization of AFLPs was also performed. Selective amplification can be done with 64 commercially available primer combinations (PC) for EcoRI and MseI restriction enzymes. Only three PCs were chosen for *Elaeis oleifera* germplasm characterization. The methodology of Vos *et al*. (1995) was used with some

modifications that reduced the time (restriction-ligation into one step) and the final volume to 11 µl.

Table 1. Information for 26 microsatellites evaluated in ASD's compact clones and *Elaeis oleifera* germplasm

Microsatellite name	EMBL acession	Motif	No. of alleles reported	Allele range expected (pb)	Annealing temperature (°C)
mEgCIR0008	AJ271625	(GA)18	11	195-220	52
mEgCIR0009	AJ271633	(GA)20	9	162-204	53
mEgCIR0018	AJ271634	(GA)18	12	158-177	52
mEgCIR0046	AJ271635	(GA)19	13	198-262	52
mEgCIR0067	AJ271636	(GA)17	13	135-187	48
mEgCIR0134	AJ271714	(GA)15	5	128-162	52
mEgCIR0219	AJ271637	(GA)17	9	205-233	52
mEgCIR0221_A	AJ271638	(GA)11	9	195-213	52
mEgCIR0221_B	AJ271638	(GA)11	0	nd	52
mEgCIR0230	AJ271639	(TA)6GAG(GA)19	11	326-354	55
mEgCIR0254	AJ271640	(GA)18	16	148-179	52
mEgCIR0304	AJ271931	(GT)4(GC)2(GT)2	3	106-114	59
mEgCIR0326	AJ271932	(GT)9	4	275-281	59
mEgCIR0332	AJ271933	(GT)8	4	131-147	52
mEgCIR0350_A	AJ271934	(GT)8	5	269-281	55
mEgCIR0350_B	AJ271934	(GT)8	0	nd	55
mEgCIR0353	AJ271935	(GT)11(GA)15	12	80-102	61
mEgCIR0377	AJ271936	(GT)6(GC)4	3	145-149	52
mEgCIR0391	AJ271938	(CCG)5	4	293-297	55
mEgCIR0437	AJ271939	(CCG)6	4	196-206	60
mEgCIR0465	AJ271940	(CCG)6	5	125-137	59
mEgCIR0476	AJ271941	(CCG)7	5	165-177	52
mEgCIR1772	AJ271937	(GT)22	15	166-198	55
mEgCIR3194	AJ578625	(GA)11	1*	266	52
mEgCIR3275	AJ578630	(GA)17	1*	146	58
mEgCIR3286	AJ578633	(GA)19	1*	121	58
mEgCIR3413	AJ578665	(GA)18	1*	223	58
mEgCIR3716	AJ578711	(GA)19	1*	200	61

EMBL accession: accession number of the European Molecular Biology Laboratory; alleles reported according to Billotte *et al.* 2001 and 2005 (*); nd: not previously reported

Visualization of fragments amplified with SSR and AFLP

Polyacrylamide gel electrophoresis (PAGE) and capillary electrophoresis in the ABI 3130 genetic analyzer (Applied Biosystems) are the systems used at ASD to visualize and score fragments amplified with microsatellites and AFLPs. Capillary electrophoresis is used mainly because results are obtained faster than PAGE and resolution is higher, which allows the detection of alleles differing in 2 base pairs with microsatellites and 1 base pair with AFLP fragments. PCR products were fluorescently marked with four

different dyes and then multiplexed to efficiently obtain fragment visualization in the ABI 3130 in less time.

There are two factors affecting the interpretation and genotyping of samples: the expected fragment size and the signal intensity of fluorescent dyes. In the multiplex, the expected fragment sizes of microsatellites shouldn't overlap in at least 20 base pairs and the intensity of fragments in relative fluorescence units (RFU) should be similar for all dyes. In the analyses performed it was found that intensities are 6-FAM>VIC>NED>PET. The dilution of dyes causes detection and analysis failures and the concentration of dyes generate peaks exceeding threshold detection. These peaks are less precise and misshapen and cause pull-ups that are detected in wrong absorbance ranges that could lead to misinterpretations.

Fingerprinting of ASD's compact clones and parentage analysis

Twenty compact ortets (donor tissues for cloning) and their clones were used to perform fingerprinting analyses (Table 2). These genotypes were analyzed with 17 SSRs primers that were selected for the easy interpretation of amplified alleles in the ABI 3130 genetic analyzer, polymorphism levels previously evaluated with several genotypes (see table 3) and reproducibility of amplified alleles in the validation process in a side by side experiment that generated two fingerprints.

During the first stage of validation, DNA extraction and amplification of microsatellites by PCR were performed in the Biotechnology Research Center laboratories at the Technological Institute of Costa Rica. The amplified fragments were sent to the Cell and Molecular Biology Research Center, University of Costa Rica, where fragments were detected in the ABI 3130 genetic analyzer (Applied Biosystems). The second step was implemented in ASD's molecular biology laboratory in Coto, Southern Costa Rica, where the same methodologies were applied, following the same protocol for DNA extraction and using the same PCR conditions for fragment amplification of microsatellites. Instead, capillary electrophoresis was conducted in ASD's ABI 3130.

The genotyping of clones with data generated by both genetic analyzers was performed with Genotyper® and GeneMapper® softwares and in all cases GS-500-LIZ® was used as the internal size standard for the allele calling and allele sizing generated by microsatellites.

Table 2. Yield performances of ASD Compact clones originated from ten different compact lines used for fingerprinting

Clone	Line	FFB	O/B
Heracles	93C351	136	32.8
Starlight	95C098	157	33.7
Sergio	C333	165	28.2
Discovery	C9235	154	31.8
Salmon	C9235	187	29.4
Titan	C9235	138	25.5
Endurance	C9236	149	32.4
Epsilon	C9236	179	33.5
Odin	C9236	179	32.9
Opal	C9236	149	32.9
Prince	C9236	178	32.4
Sabre	C9236	114	30.3
Sunrise	C9236	184	29.2
Thor	C9236	159	33.6
Petit	C9259	204	29.3
Eagle	C9268	158	27.3
Ares	C9269	199	33.7
Alpha	C9274	189	26.3
Tornado	C9274	174	31.6
Fran	C9278	207	28.4
Average		168	30.8

FFB= fresh fruit bunch, kg/palm/year, six years average; O/B oil to bunch (%), average of eight bunch analyses per palm

Variability and genetic relationship in ASD's *Elaeis oleifera* germplasm

The analysis of genetic relationships was done on *Elaeis oleifera* germplasm from the first and second generations (see Table 5) that was originated from populations introduced to Costa Rica during the 1960s and 1970s (Escobar 1981). Eight SSR primers previously described by Billotte *et al*. (2001) were used and amplification products were separated in an ABI 3130 genetic analyzer (Applied Biosystems) by adding 0.5 µL of internal size standard GS-500-LIZ® for allele sizing.

The germplasm of *E. oleifera* was also analyzed with three primer combinations for EcoRI/MseI enzymes: AGA x CAC, AGA x CAG and AAC x CAC. The amplified fragments were separated in the ABI 3130 genetic analyzer and fragment calling and sizing were performed with GeneMapper® software by using GS-500-ROX®. Despite

establishing a baseline of 50 RFU, only those fragments with signal intensity greater than 75 RFU were included in the analysis, considering that fragments close to the baseline might cause distortion in data interpretation. In addition, better separation of genotype groups is obtained in multidimensional scaling analyses (Araya *et al.* 2005a).

RESULTS AND DISCUSSION

DNA isolation

DNA isolated from adult palm leaflets had high quality. Electrophoresis on 1% agarose gel (figure 1, left panel) shows the DNA obtained with this method. DNA was of high quality and approximately 50-100 ng/uL was isolated with 75 mg of leaflet tissue. When this method was used on young plant material, DNA quality was low and showed the presence of contaminants that were not removed during the extraction (Figure 1, center panel). This result led to the modification of the extraction method for young palm leaflets.

Phenolic extraction was performed to denature proteins that were not completely removed. The cleaning steps with chloroform:octanol resulted in the solubilization of phenol and removal of proteins, yielding high quality DNA (Figure 1, right panel). The extraction buffers used in the two methods contained both PVP (polyvinylpyrrolidone) and CTAB to reduce problems caused by the presence of phenolic compounds and polysaccharides; protein removal helped prevent degradation of the DNA (Valadez and Kahl, 2000).

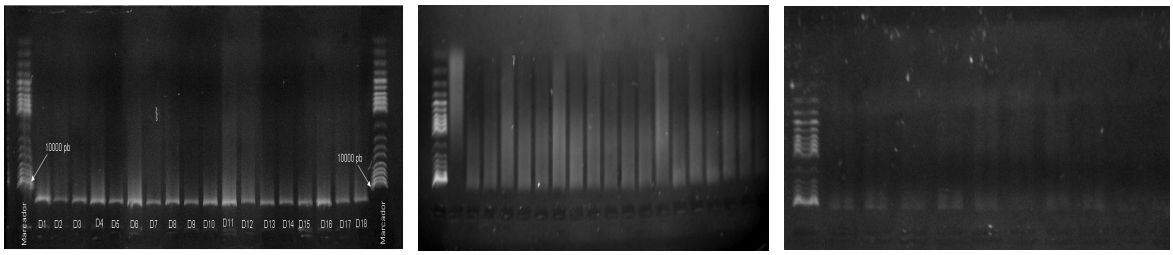

Figure 1**.** DNA quality obtained from oil palm leaflets. **Left** p**anel:** DNA from 15 year-old palms obtained using the method for adult plant material. **Central panel:** DNA from nursery stage palms obtained using the method for adult plant material. **Right panel:** DNA from nursery stage palms obtained using the method for young plant material.

Microsatellite analyses

Standardization of the DNA extraction method and evaluation of PCR parameters for microsatellite amplification were the first steps in research involving the analysis of compact clones and *Elaeis oleifera* germplasm.

Twenty six primers were evaluated in the optimization phase (Table 1) but only 19 were kept for further analysis because some exhibited amplification problems showing unclear banding patterns and difficulty with allele determination. These 19 microsatellites showed an average observed heterozygosity of 0.62 and a probability of identity of 0.25

when combining data from studies on ASD *E. guineensis* and *E. oleifera* germplasm (Table 3).

One of the parameters shown in Table 3 is the probability of identity, which provides an estimate of the probability that two unrelated individuals have the same genotype in a population with random mating (Peakall and Smouse, 2006). The probability of identity is used as an indicator of statistical power provided by microsatellites. Thus, primers such as mEgCIR0353, mEgCIR3286 and mEgCIR3716 have lower individual probabilities that coincide with the high number of alleles, mainly in mEgCIR0353. Although some primers had high probabilities (mEgCIR0437, mEgCIR0332 and mEgCIR0021) the product of the overall probabilities was very low (2.24×10^{-15}), which means that these microsatellites are suitable for use in applications such as genetic fingerprinting.

Table 3. Genetic information parameters of 19 microsatellites used to evaluate *Elaeis* sp. germplasm of ASD.

Microsatellite name	Species	N	Allele range observed (bp)	NA	Ho	PI
mEgCIR0008	Eg & Eo	210	196-222	11	0.79	0.20
mEgCIR0009	Eg & Eo	210	160-179	8	0.75	0.21
mEgCIR0046	Eg & Eo	210	200-239	20	0.90	0.16
mEgCIR0067	Eg	18	147-160	5	0.72	0.12
mEgCIR0219	Eg	18	205-236	9	0.94	0.18
mEgCIR0221_A	Eg	18	197-210	3	0.39	0.47
mEgCIR0221_B	Eg	18	223-233	4	0.59	0.20
mEgCIR0254	Eg	20	153-167	5	0.80	0.11
mEgCIR0304	Eo	190	106-116	4	0.98	0.31
mEgCIR0332	Eg	18	145-150	2	0.50	0.48
mEgCIR0350_A	Eg	18	274-285	4	0.22	0.41
mEgCIR0350_B	Eg	18	301-311	4	0.44	0.29
mEgCIR0353	Eg & Eo	210	80-105	11	0.57	0.03
mEgCIR0391	Eg	18	289-298	6	0.55	0.26
mEgCIR0437	Eg & Eo	210	195-209	4	0.53	0.60
mEgCIR0465	Eg & Eo	210	123-132	4	0.35	0.40
mEgCIR1772	Eg & Eo	210	165-185	8	0.52	0.17
mEgCIR3275	Eg	48	127-140	3	0.61	0.30
mEgCIR3286	Eg	48	117-124	6	0.76	0.07
mEgCIR3413	Eg	48	210-230	6	0.55	0.14
mEgCIR3716	Eg	48	186-206	6	0.55	0.09
Average		96		6	0.62	0.25

N: number of samples; NA: Number of alleles observed in both species;
Ho: combined observed heterozygosity; PI: probability of identity

AFLP analyses

DNA samples were digested and ligated with the optimized method and a visible smear was observed in the 100-1500 bp range, suggesting that ligation of the adapters and preselective amplification were effective. The digestion of genomic DNA with EcoRI and MseI enzymes results in three fragment types: MseI-MseI, EcoRI-EcoRI and EcoRI-

MseI. It is expected that over 90% of the fragments are MseI-MseI, a number of EcoRI-MseI fragments are twofold EcoRI restriction sites and that only a small proportion are EcoRI fragments (Vos *et al.* 1995). When the concentration of MseI was decreased in the restriction-ligation mixture and EcoRI remained constant, the proportion of MseI-MseI fragments were reduced and increased the EcoRI-MseI fragments, resulting in a more effective method of adapters ligation.

Fingerprinting of ASD's compact clones and parentage analysis

In the validation process two identical fingerprintings of clones were obtained in the side by side work performed in two different laboratories and two ABI 3130 genetic analyzers. This result suggested that the protocol for molecular identification of oil palm genotypes is reliable and reproducible. The results revealed that seventeen primers amplified a total of 81 alleles with an average of 4.05 ± 1.39 alleles per locus and an expected heterozygosity (He) ranging from 0.05 to 0.80 (Table 4).

Of these 17 microsatellites, mEgCIR0067, mEgCIR0254, mEgCIR1772 exhibited the highest heterozygosity (> 0.70), which coincides with the results described by Billotte *et al.* (2001). Three primers (mEgCIR0219, mEgCIR0350 and mEgCIR0221) showed double putative amplification of SSRs, i.e. there is a duplication of these loci in the genome of the individuals studied, which has not been previously described.

Table 4. Content of genetic information parameters of 17 microsatellites, estimated on 20 compact clones of ASD

Microsatellite	Na	I	He	PI	PE
mEgCIR0046	3	0.91	0.54	0.27	0.37
mEgCIR0067	5	1.38	0.72	0.13	0.60
mEgCIR3286	5	1.54	0.78	0.09	0.60
mEgCIR0254	7	1.54	0.75	0.11	0.71
mEgCIR0350_A	4	0.70	0.34	0.45	0.50
mEgCIR0350_B	3	0.85	0.49	0.31	0.37
mEgCIR3413	5	1.05	0.56	0.25	0.60
mEgCIR3716	5	1.30	0.69	0.15	0.60
mEgCIR0008	3	0.74	0.41	0.39	0.37
mEgCIR0332	2	0.61	0.42	0.43	0.19
mEgCIR1772	6	1.49	0.74	0.11	0.66
mEgCIR0009	3	1.07	0.65	0.20	0.37
mEgCIR0219_A	4	1.18	0.66	0.18	0.50
mEgCIR0219_B	2	0.12	0.05	0.91	0.19
mEgCIR0353	6	1.65	0.80	0.07	0.66
mEgCIR0465	3	0.65	0.37	0.44	0.37
mEgCIR0221_A	4	0.68	0.34	0.46	0.50
mEgCIR0221_B	5	1.23	0.67	0.17	0.60
mEgCIR0391	3	0.70	0.43	0.40	0.37
mEgCIR3275	3	0.78	0.52	0.34	0.37
Total	81			3.60×10^{-13}*	0.99999**
Average	4.05	1.01	0.55	0.29	0.47

Na: number of observed alleles; I: information index; He: expected heterozygosity; PI: probability of identity; PE: probability of exclusion; * product of PIs; ** 1-(product of PEs)

Allele frequencies were used to calculate the probability of identity (PI) using GenAlEx (Peakall and Smouse, 2006), a Microsoft® Excel add-in. The PI expresses the probability of finding two individuals with the same genotype in a given locus in the individuals analyzed. Considering that each locus segregates independently, the probability of finding identical genotypes is the product of the probabilities of the 17 loci, which were in a range between 0.07 and 0.91 (Table 4). The combined PI was 3.60×10^{-13} which means that finding two individuals with the same genotype is an almost null event.

Besides PI, the probability of exclusion indicates the probability that a male or female parent is excluded in a paternity test. The values obtained for this parameter ranged from 0.19 to 0.71 (Table 4). The combined power of exclusion for the 17 microsatellites was > 0.99999, indicating that the set of loci evaluated is suitable for parentage analysis of clones, verification of contamination and assessing its legitimacy.

For the parentage analysis, pairwise genetic distances were calculated clone by clone, generating a distance matrix of N x N individuals with a total of 200 comparisons. This matrix had values >1 because the analysis considers the quadratic distances between individuals with i-th, j-th, k-th, l-th different alleles in one locus. The average genetic distance between clones was 16.64 with a range of 4.73 to 24.02.

The distribution of frequencies of genetic distances resulted in 168 comparisons above the average distance, meaning that 84% of genetic distances are medium to high and suitable for the parentage analysis. For example, the lowest scored genetic distance (4.73) was between the clones named Eagle and Salmon, which are half-sibs, suggesting that a high proportion of their genetic elements come from their common progenitor. The pairwise comparisons of some full-sib clones scored low values (close to 4.73), suggesting a high degree of parentage. Principal Coordinates Analysis (Figure 2) accounted for 46.3% of the observed variation and showed the grouping of clones from the same family (line) members (C9236, table 2). The similarities between Odin, Opal, Sabre, Endurance and Prince are highlighted; also similarities between Thor, Sunrise and Epsilon were observed. Although Starlight, Heracles and Ares clones belong to different families, they share 50% of the genetic background with C9236 family, which explains their relatedness in the PCoA.

These results demonstrated the suitability of microsatellite markers for molecular identification of oil palm compact clones by using a reliable and high discrimination set of microsatellites. These markers were also suitable for assessing the variation and genetic distance between clones and these applications might be extended to study breeding populations.

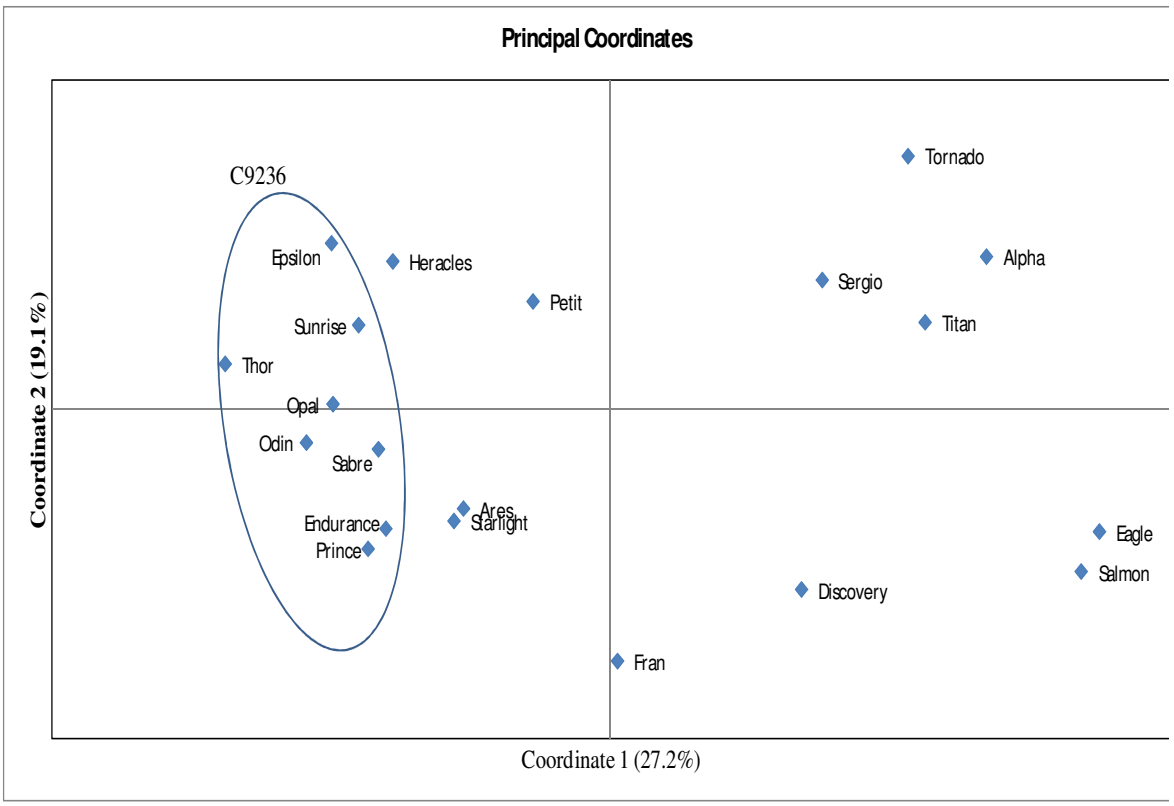

Figure 2. Principal Coordinates Analysis based on pairwise genetic distances of ASD compact clones.

Variability and genetic relationships between ASD *Elaeis oleifera* germplasm

Microsatellite analysis

The microsatellites used were highly informative for analysis of genetic relationships in *Elaeis oleifera* germplasm with polymorphism levels higher than 75% (Table 5) across the populations studied. Some exhibited 100% polymorphism as in F1 Brazil, F1 Colombia, F1 Nicaragua x Honduras and F2 Costa Rica/Colombia x Costa Rica/Panama. The high expected heterozygosity (He) scored in all microsatellites resulted in high levels of polymorphism. Most of the SSRs had He above 0.54 with remarkable values of 0.90 and 0.77 for mEgCIR0046 and mEgCIR0008, respectively (Table 6).

Table 5. Genetic information parameters obtained with 8 SSRs in the *Elaeis oleifera* germplasm of ASD

Population	No. of samples	Observed alleles	Allele per locus	Ho	He	Polymorphic loci (%)
Bra x CR/Bra	10	33	4.13	0.52	0.54	88
Bra x CR F1	10	28	3.50	0.80	0.53	88
Ecuador OP	9	22	2.75	0.33	0.32	50
F1 Bra	11	39	4.88	0.46	0.53	100
F1 Bra x CR	2	17	2.13	0.56	0.38	75
F1 Col	10	23	2.88	0.34	0.36	100
F1 CR	11	21	2.63	0.35	0.32	75
F1 CR x Col	10	23	2.88	0.41	0.34	75
F1 CR x Hon	9	23	2.88	0.48	0.40	88
F1 CR x Nic	10	23	2.88	0.49	0.40	88
F1 CR x Pan	10	28	3.50	0.44	0.46	88
F1 Nic x Hon	10	25	3.13	0.34	0.43	100
F1 Nic x Pan	9	24	3.00	0.52	0.45	88
F1 Pan	8	24	3.00	0.34	0.41	88
F1 Pan x Col	9	25	3.13	0.47	0.40	88
F1 Pan x Hon	10	23	2.88	0.56	0.44	88
F2 CR/Bra	3	18	2.25	0.42	0.47	88
F2 CR/Bra x CR	2	16	2.00	0.44	0.36	63
F2 CR/Col x CR/Pan	9	30	3.75	0.35	0.45	100
F2 CR/Pan x CR	4	16	2.00	0.38	0.34	63
F2 CR/Pan x CR/Col	8	19	2.38	0.36	0.32	75
F2 CR/Pan x CR/Pan	6	22	2.75	0.56	0.41	75
Surinam	10	15	1.88	0.15	0.16	50
Total	190					
Average	8.26	23.35	2.92	0.44	0.40	81

Bra=Brazil; Col=Colombia; CR=Costa Rica; Hon=Honduras; Nic=Nicaragua; OP=open pollination; Pan=Panama

Table 6. Number of observed alleles (Na), effective number of alleles (Ne), observed heterozygosity (Ho) and expected heterozygosity obtained with 8 loci in the *Elaeis oleifera* germplasm of ASD.

Locus	Na	Na*	Ne	Ho	He	He*
mEgCIR0353	8	6	2.66	0.58	0.62	0.70
mEgCIR0009	8	7	2.40	0.26	0.58	0.82
mEgCIR0437	4	3	2.58	0.54	0.61	0.55
mEgCIR0465	3	3	2.18	0.32	0.54	0.54
mEgCIR1772	8	8	2.12	0.27	0.53	0.84
mEgCIR0008	11	11	4.37	0.58	0.77	0.87
mEgCIR0304	4	3	1.99	0.14	0.50	0.22
mEgCIR0046	20	11	10.20	0.79	0.90	0.86

Na* and He*: number of alleles and expected heterozygosity described by Billote *et al*. 2001

Allele frequencies were used to calculate the genetic distances (Nei) between the germplasm of *E. oleifera* and Principal Coordinates Analysis was also performed. The first two axes in the PCoA explained 80.6% of molecular variation between the origins analyzed (Figure 3). The germplasm was separated according to geographical regions: Brazil, Surinam and Central America/Colombia, which is consistent with the groups described by Billotte *et al*. (2001) and Barcelos *et al.* (2002); the germplasm from Ecuador was located near the Brazilian ones.

Higher levels of genetic diversity were obtained in the group conformed by the F1 Brazil population and the progenies derived from it, which yielded an expected heterozygosity > 0.47 (Table 5) and clearly formed a group in the PCoA. The F1 Brazil x Costa Rica had a lower level of heterozygosity; however, in the PCoA it remained in the same group. The germplasm of Ecuador is not as diverse (He = 0.33) but it shares a high percentage of genetic elements with those materials from Brazil (Figure 3). The separation from the Central American group resulted of the exclusive amplification of alleles at loci mEgCIR1772, mEgCIR0008 and mEgCIR0009 within Brazil and Ecuador germplasm (Table 7).

The lowest genetic diversity was observed in the germplasm of Suriname (He = 0.15) that also showed the greatest genetic divergence with the remaining origins. The exclusive amplification of alleles at loci mEgCIR0009 and mEgCIR0465 and the presence of an allele with a frequency of 70% for loci mEgCIR0353 caused the wide separation observed in the Suriname population (Table 7). On the other side, the expected heterozygosity showed intermediate genetic diversity ranging from 0.32 to 0.46 in the germplasm of Central America/Colombia. The low genetic divergence in these groups was revealed in a close-related group of the PCoA without distinction between Central America, Panama and Colombia origins (Figure 3). This suggests that *E. oleifera* materials from Central America in the first and second generation might have fixed a high percentage of their genetic elements and show low to intermediate heterozygosity as revealed with microsatellites.

In summary, the genetic diversity of the groups was Brazil>Central America/Colombia>Surinam. Barcelos *et al.* (2002) described similar results in the populations studied, with very low levels of genetic diversity in the population of Suriname and the highest diversity index in the materials from Brazil.

The microsatellite analyses of ASD Costa Rica *E. oleifera* germplasm suggested low levels of genetic divergence within the Central American and Colombian materials, plus greater divergence from those of Brazil. This information may be considered in the future for designing breeding programs in two ways: to increase the heterozygosity through the combination of genetically divergent populations, for example Brazil and Ecuador to Central America, where the former has the highest oil content in the mesocarp (> 28%) and those of Ecuador possess the greatest mesocarp to fruit ratio (> 64%). Conversely, allele fixation to increase homozygosis might be reached by intercrossing Central America germplasm that is remarkable for its good fruit set, showing fertile fruits above 46% (Alvarado *et al.* 2009).

These results show the usefulness of microsatellites for measuring genetic diversity between *oleifera* origins; moreover, the transferability of these markers will allow their use in evaluating *E. guineensis* materials. The application of SSRs can be expanded in controlled crossing programs to measure the degree of introgression, especially in creating new hybrids between E. *oleifera* and *E. guineensis.*

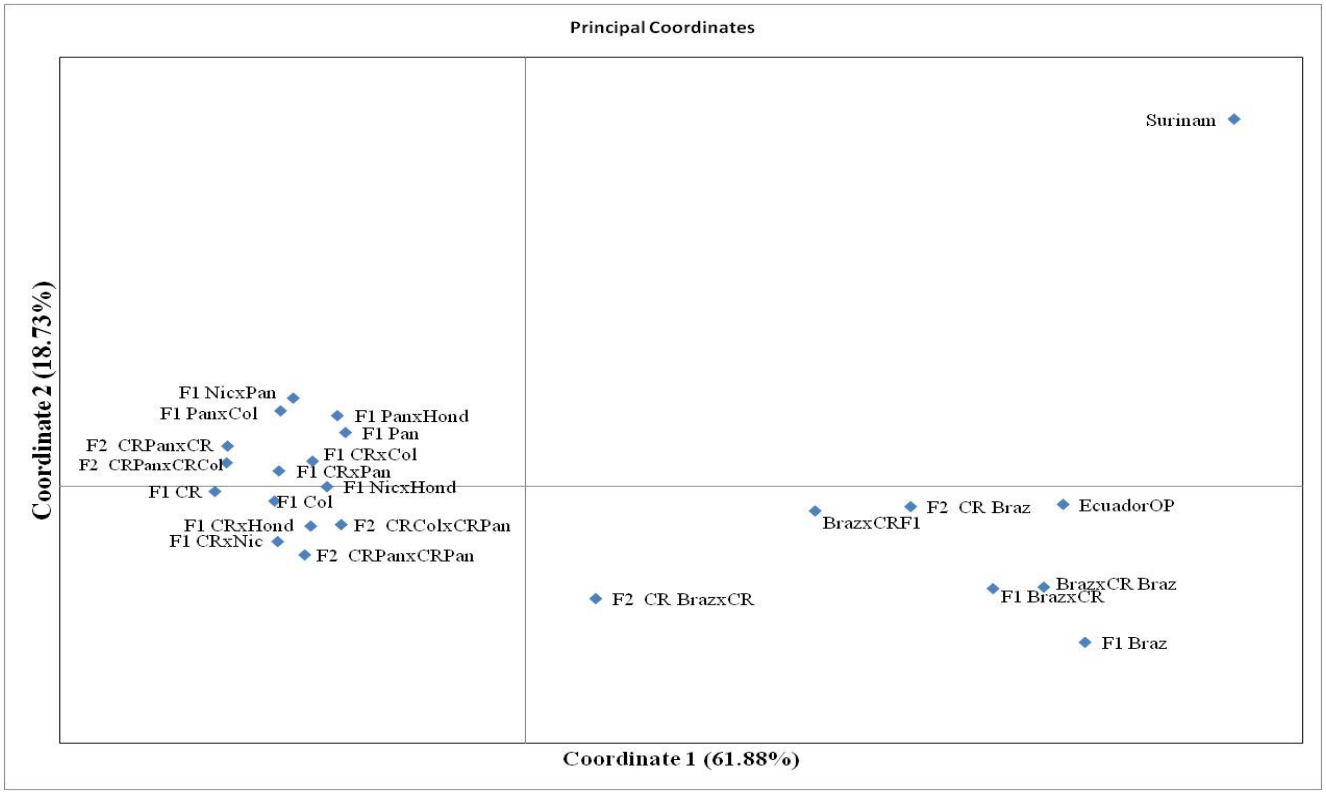

Figure 3. Principal Coordinates Analysis of Elaeis oleifera germplasm amplified with 8 microsatellites

Table 7. Exclusive allele amplified in the *Elaeis oleifera* germplasm of ASD

Origin	Locus	Allele (bp)	Frequency
Brazil x Costa Rica Brazil	mEgCIR0009	166	0.11
Ecuador OP	mEgCIR1772	185	0.67
F1 Brazil	mEgCIR1772	165	0.05
F1 Brazil	mEgCIR0008	207	0.05
F1 NicaraguaxHonduras	mEgCIR1772	182	0.06
F1 NicaraguaxPanama	mEgCIR0008	211	0.22
F2 Costa Rica Brazil	mEgCIR0009	175	0.17
Surinam	mEgCIR0353	85	0.70
Surinam	mEgCIR0009	160	1.00
Surinam	mEgCIR0465	123	1.00

AFLP analyses

The analysis of *E. oleifera* with primer combinations (PC) AGA x CAC, AGA x CAG and AAC x CAC produced 58, 113 and 98 fragments respectively, which showed a frequency ≥ 5% across all populations studied. The registered polymorphism was lower when compared to microsatellites, ranging between 21.19 and 65.43% (Table 8). F1 Panama x Colombia, F1 Brazil and F2 Costa Rica/Colombia x Costa Rica/Panama populations were the most polymorphic with values > 60% (Table 8). Although AFLPs have the advantage of covering a larger proportion of the genome, the heterozygosity across all populations was very narrow (0.09-0.20) with an average of 0.15 ± 0.03. Some populations that showed middle to high heterozygosity with microsatellites did not necessarily show the same level with AFLPs. This happened with F1 Brazil x Costa Rica, F2 Costa Rica/Panama x Costa Rica, F2 Costa Rica/Brazil x Costa Rica and F2 Costa Rica/Brazil populations; few samples were analyzed from these groups, possibly leading to the lowest heterozygosity values (< 0.11) scored (Table 8).

Consistently with microsatellites, the F1 Brazil population had one of the highest heterozygosity levels (0.19) revealed by AFLP. This value is lower than the heterozygosity obtained with microsatellites (0.53, Table 5) but it was of the same magnitude when compared with all the origins. The lowest heterozygosity (0.11) was observed in Suriname germplasm despite the number of samples analyzed (n = 10, Table 8).

Polymorphic fragments were used in a PCoA where the first two axes accounted for 72.78% of the molecular variation (Figure 4). In this plane and consistent with microsatellites, germplasm from Brazil and derivative progenies were kept separate from other populations. Ecuador germplasm were separated with AFLPs that were in the same group as the Brazilian material when microsatellites were used (Figure 3). By contrast, the Suriname population is not so widely separated from the rest of the germplasm as occurred with microsatellites (Figure 3) because the magnitude of divergence was

smaller. The AFLPs were able to distinguish three groups within the origins of Central America/Colombia. These were composed of i) the F2 populations with genetic elements of Costa Rica, Panama and Colombia, ii) the F1 populations originating from Central America and iii) the F1 population of Colombia and the crossing between Colombia and Costa Rica. In the first group the lowest divergence was observed with a genetic distance < 0.060 between populations, possibly because these materials represent an advanced breeding population and thus the genetic base is reduced. In the second group the proximity of F1 Costa Rica x Honduras with F1 Costa Rica x Nicaragua resulted from the lowest genetic distance (0.015) scored in all pairwise comparisons. Moreover, these two sources are separated slightly from the group, leaving aside those with genetic elements of Panama and whose subgroup showed a very low average divergence (0.064). Finally, the third group showed the F1 Colombia population to be closely grouped with the F1 Costa Rica x Colombia with a distance of 0.034.

Table 8. Genetic diversity parameters estimated with 269 AFLP markers amplified across 23 *Elaeis oleifera* origins.

Population	No. of samples	No. of fragments	Het.	Polymorphism (%)
Braz x CR/Braz	10	171	0.16	58.36
Braz x CR F1	10	136	0.16	50.19
Ecuador OP	9	130	0.15	47.96
F1 Braz	11	184	0.19	63.94
F1 Braz x CR	2	100	0.09	21.19
F1 Col	10	119	0.12	39.78
F1 CR	11	162	0.16	52.79
F1 CR x Col	10	158	0.15	53.16
F1 CR x Hon	9	151	0.15	50.93
F1 CR x Nic	10	117	0.12	36.43
F1 CR x Pan	10	169	0.17	58.36
F1 Nic x Hon	10	131	0.14	48.33
F1 Nic x Pan	9	170	0.18	58.74
F1 Pan	8	147	0.17	52.79
F1 Pan x Col	9	188	0.19	62.83
F1 Pan x Hon	10	162	0.19	60.22
F2 CR/Braz	3	113	0.11	30.48
F2 CR/Braz x CR	2	104	0.09	22.68
F2 CR/Col x CR/Pan	9	180	0.20	65.43
F2 CR/Pan x CR	4	117	0.09	25.65
F2 CR/Pan x CR/Col	8	181	0.18	60.97
F2 CR/Pan x CR/Pan	6	161	0.16	49.81
Surinam	10	117	0.11	37.55
Average	8.26	146.43	0.15	48.20

Het= heterozygosity; Braz=Brazil; Col=Colombia; CR=Costa Rica; Hon=Honduras; Nic=Nicaragua; PA=open pollination; Pan=Panama

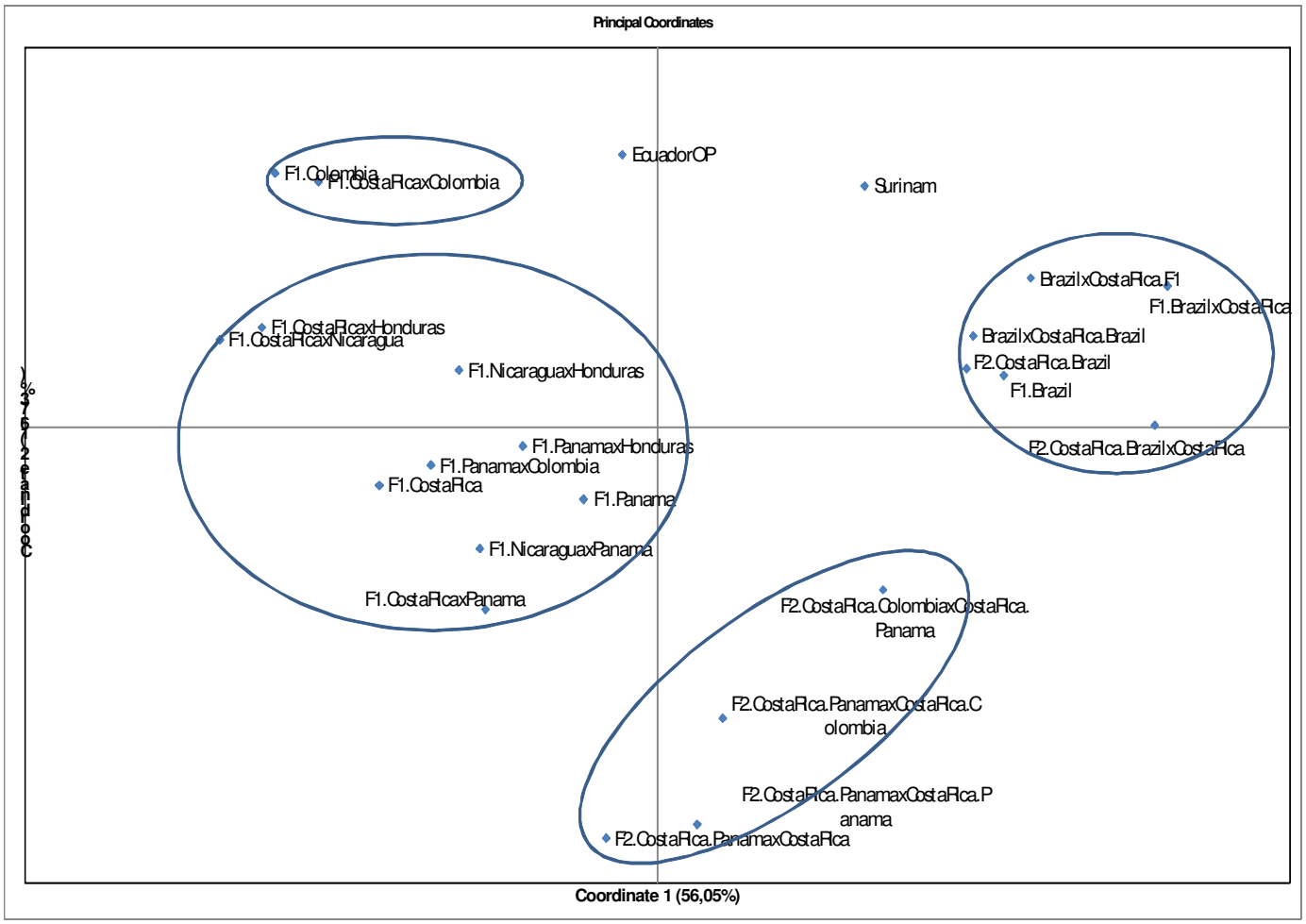

Figure 4**.** Principal Coordinate Analysis of ASD *Elaeis oleifera* germplasm revealed with 269 AFLP markers.

CONCLUDING REMARKS

The genetic diversity and relationships between *Elaeis guineensis* and *Elaeis oleifera* germplasm will continue to be evaluated with the aim of supporting the ASD breeding program. This information will aid breeding decisions because it might allow the creation of new populations seeking greater genetic variation through the recombination of distant individuals or the concentration of favorable traits.

Microsatellite markers proved to be precise and reliable with a high capacity of discrimination and thus suitable for molecular identification of oil palm genotypes. Because of these characteristics of SSRs coupled with the advantages of AFLPs (wider genome coverage and thus a higher number of markers) ASD is developing a linkage map in a compact segregating population characterized by low trunk height increment and short leaf length. The aim is to develop marker assisted selection of parental lines for

the generation of new varieties. At an early stage, the use of Bulked Segregant Analysis allowed the detection of SSR markers associated exclusively with compact germplasm.

AFLPs and microsatellites are useful tools for finding markers linked to or associated with the compact trait in the oil palm genome; however, this relation may be lost in more advanced compact generations. This limitation can be overcome using the tools of functional genomics to gain a deeper understanding of the genes involved in determining the compact trait. This also requires a better understanding of the anatomical differences between compact and conventional genotypes. This knowledge will be the basis for generating cDNA libraries and sequences that will serve gene annotation. Candidate genes can be integrated into the reference genetic maps generated by molecular markers, in order to have more elements for assisted selection using molecular biology tools.

LITERATURE CITED

Agüero, L. 2007. Optimización y validación de la técnica de marcadores microsatelitales en palma aceitera (*Elaeis guineensis* Jacquin). Trabajo Final de Graduación BSc. Cartago, Instituto Tecnológico de Costa Rica. 62 p.

Alvarado, A; Escobar, R; Peralta, F. 2009. Avances en el mejoramiento genético de la palma de aceite en Centro América. *In* XVI International Oil Palm Conference. Cartagena, Colombia.

Araya, E; Murillo, O; Aguilar, G; Rocha, O. 2005a. Relaciones genéticas en una colección de clones de Gmelina arborea (Roxb) reveladas con marcadores AFLP. Kuru: Revista Forestal. 2(6):1-14.

Araya, E; Murillo, O; Aguilar, G; Rocha, O. 2005b. A DNA extraction and initial primers screening in Hyeronima alchorneoides Fr. all for AFLP applications. Foresta Veracruzana 7(1): 1-4.

Bakoume, C; Wickneswari, R; Rajanaidu, N; Kushairi, A; Amblard, P; Billotte, N. 2006. Allelic diversity of natural oil palm (*Elaeis guineensis* Jacq.) populations detected by microsatellite markers. Implication in conservation. In XV Conferencia Internacional sobre Palma de Aceite, 19-22 de setiembre, Cartagena de Indias, Colombia. 26 p.

Barcelos, E; Amblard, P; Berthaud, J; Seguin, M. 2002. Genetic diversity and relationship in American and African oil palm as revealed by RFLP and AFLP molecular markers. Pesquisa Agropecuaria Brasileira 37(8):1105-1114.

Billotte, N; Risterucci, AM; Barcelos, E; Noyer JL, Amblard, P; Baurens FC. 2001. Development, characterization, and across-taxa utility of oil palm (*Elaeis guineensis* Jacq) microsatellite markers. Genome 44:413-425.

Billotte, N; Marseillac, N; Risterucci, AM, Adon, B; Brottier, P; Baurens, FC; Singh, R; Herran, A; Asmady, H; Billot, C; Amblard, P; Durand-Gasselin, T; Courtois, B; Asmono, D; Cheah, SC; Rohde, W; Ritter, E; Charrier, A. 2005. Microsatellite-based high density linkage map in oil palm (*Elaeis guineensis* Jacq.). Theoretical and Applied Genetics 110:754-765.

Escobar, R. 1981. Preliminary results of the collection and evaluation of the American oil palm (*Elaeis oleifera* HBK Cortes) in Costa Rica. Proc. Int. Conf. on Oil Palm

in Agriculture in the Eighties. Kuala Lumpur, 17-20 June, 1981.The Incorporated Society of Planters, pp 79-97.

Fernández, M. 2008. Diferenciación de los tipos de fruto de palma africana (*Elaeis guineensis* Jacq.) mediante marcadores moleculares: AFLP y SSR. Trabajo Final de Graduación BSc. Cartago, CR, Instituto Tecnológico de Costa Rica. 71 p.

Hayati, A; Wickneswari, R; Maizura, I; Rajanaidu, N. 2004. Genetic diversity of oil palm (*Elaeis guineensis* Jacq.) germplasm collections from Africa: implications for improvement and conservation of genetic resources. Theor. Appl. Genet. 108(7):1274-84.

Jack, PI; James, C; Price, Z; Rance, K; Gorvers, L; Corley, RHV; Nelson, S; Rao, V. 1998. Application of DNA markers in oil palm breeding. In International Oil Palm Conference. 23-25 setiembre. Nusa Dua, Bali, MY.

Mayes, S; Jack, PL; Corley, RHV. 2000. The use of molecular markers to investigate the genetic structure of an oil palm breeding programme. Heredity 85:288-293.

Moretzsohn, MC; Ferreira, MA; Amaral, ZPS; Coelho, PJA; Grattapaglia, D; Ferreira, ME. 2002. Genetic diversity of Brazilian oil pal (*Elaeis oleifera* H.B.K.) germplasm collected in the Amazon Forest. Euphytica 124:35-45

Norziha, A; Rafii, YM; Maizura, I; Ghizan, S. 2007. Genetic variation among oil palm parental genotypes and their progenies based on microsatellite markers. In Proceedings of the PIPOC 2007 International Palm Oil Congress (Agriculture, Biotechnology and Sustainability), 26-30 de agosto, Kuala Lumpur, Malaysia. pp 940-953.

Peakall, R; Smouse, PE. 2006. GENALEX 6: genetic analysis in Excel. Population genetic software for teaching and research. Molecular Ecology Notes. 6:288-295.

Purba, AR; Noyer, JL; Baudouin, L; Perrier, X; Hamon, S; Lagoda, PJL. 2000. A new aspect of genetic diversity of Indonesian oil palm (*Elaeis guineensis* Jacq.) revealed by isoenzyme and AFLP markers and its consequences for breeding. Theor. Appl. Genet. 101:956–961.

Rajanaidu, N; Jalani, S; Rao, V; Kushairi, A. 1993. Genotype-environment interaction (GE) studies in oil palm (*Elaeis guineensis*) progenies. *In* Proceedings of the 1991 ISOPB International workshop on genotype-environment interaction studies in perennial tree crops. pp 12-32, Palm Oil Research Institute, Malaysia, Kuala Lumpur.

Rocha, PJ. 2002. Teoría y práctica para la extracción y purificación del ADN de palma de aceite. Palmas 23(3):9-17.

Rocha, PJ; Rey, L. 2007. Partial agronomical, biochemical and molecular characterizations of Colombian *Elaeis oleifera* germplasm bank. In Proceedings of the PIPOC 2007 International Palm Oil Congress (Agriculture, Biotechnology and Sustainability), 26-30 de agosto, Kuala Lumpur, Malaysia. pp 954.

Singh, R; Nagappan, J; Tan, SG; Panandam, JM; Cheah, SC. 2007. Development of simple sequence repeat (SSR) markers for oil palm and their application in genetic mapping and fingerprinting of tissue culture clones. Asian Pacific Journal of Molecular Biology and Biotechnology 15(3):121-131.

Shah, FH; Rashid, O; Simona, AJ; Dunsdon, A. 1994. The utility of RAPD markers for the determination of genetic variation in oil palm (*Elaeis guineensis*). Theoretical and Applied Genetics 89:713-718.

Shah, FH; Nyuk, LM. 1996. Use of microsatellites in the determination of genetic variation and genetic relationship between various oil palm populations. In Proceedings of the 1996 PORIM International Palm Oil Congress (Agriculture), 23-28 de setiembre, Kuala Lumpur, Malaysia. p 568-574.

Sterling, F; Richardson, Dl; Alvarado, A; Montoya, C; Chaves, C. 1999. Performance of OxG *E. oleifera* Central American and Colombian biotype x *E. guineensis* interspecific hybrids. Proc. of the seminar on worldwide performance of DxP oil palm planting materials. clones and interspecific hybrids. Ed by Rajanaidu N and Jalani BS. Palm Oil Research Institute of Malaysia. Pp. 114-127.

Valadez, E; Kahl, G. 2000. Huellas de ADN en genomas de plantas: teoría y protocolos de laboratorio. Distrito Federal, ME, Mundi-Prensa. 147 p.

Vos, P; Hogers, R; Bleeker, M; Reijans, M; Van de Lee, T; Hornes, M; Frijters, A; Pot, J; Peleman, J; Kuiper, M; Zabeau, M. 1995. AFLP: a new technique for DNA fingerprinting. Nucleic Acids Research 23(21):4407-4414.