Molecular characterization of oil palm *Elaeis guineensis* Jacq. of different origins for their utilization in breeding programmes

Diana Arias[1], Iván Ochoa[2], Fernando Castro[2] and Hernán Romero[1,3]*

[1]*Oil Palm Biology and Breeding Research Program, Colombian Oil Palm Research Center – Cenipalma, Street 21, No. 42-55, Bogota, Colombia, [2]Breeding and Seed Production Program of Unipalma S.A. Cumaral, Meta, Hacienda Santa Barbara, Bogota, Colombia and [3]Department of Biology, Faculty of Science, Universidad Nacional de Colombia, Bogota, Colombia*

Abstract

Sustainable development of a breeding programme depends on having sufficient genetic variability to achieve genetic gains in each selection cycle. The aim of this study was to molecularly characterize families of the oil palm, *Elaeis guineensis* Jacq., of different origins using microsatellite molecular markers. The value of the observed heterozygosity was higher than that of the expected heterozygosity in all of the progenies. The coefficients ($G_{ST} = 0.207$ and $F_{ST} = 0.174$) and AMOVA showed genotypic differences among the evaluated families. Likewise, this was reflected in the groups obtained by the dendrogram and principal coordinate analyses. This difference could have evolved due to the enrichment of some of the families with germplasm from different origins. Therefore, genetic relationships estimated from molecular data would be convenient to select families more distant from each group and palms more distant from each family selected to reserve genetic variability. This information will guide us in the decision-making process when planning breeding programmes focused on crosses to develop new populations with an acceptable broad genetic base and adaptability. In this way, sources of resistance to biotic and abiotic factors can be identified for the development of new varieties with competitive advantages for the sector.

Keywords: expected heterozygosity; genetic differentiation; microsatellite; molecular data

Introduction

The oil palm species *Elaeis guineensis* Jacq. originates along the coast of the Gulf of Guinea in Western and Central Africa, but its exact site of origin in Central Africa has not been pinpointed. Some authors have suggested that the centre of origin is Nigeria or Cameroon (Corley and Tinker, 2003). The expansion of oil palm cultivation began in Africa and South-East Asia during the colonial period due to its multiple industrial and food uses (Corley and Tinker, 2003). Since that time, oil palm has become one of the world's main sources of vegetable oils, with a production of 43.6 million tons of oil in 2010, accounting for 26.65% of the world's production of edible oils (FAOSTAT, 2010). Most of the commercially produced oil palm seed is *tenera* material derived from the cross Deli *dura* × AVROS *pisifera*. However, the low level of diversity in the populations of Deli *dura* (Kularatne *et al.*, 2001) could limit oil productivity (Dumortier, 2003). The growing global demand for palm oil has created the need to incorporate new sources of African genes from different origins. In this way, variability can be increased in order to develop palms

* Corresponding author. E-mail: hromero@cenipalma.org

with superior agronomic qualities (Corley and Tinker, 2003). The majority of seed production programmes are based on the selection of *dura* (thick-shelled) genotypes as the female parent and *pisifera* (shell-less) genotypes as the pollen parent. This selection is performed through the evaluation of the selected breeding parents using progeny testing of crosses involving parents *dura* and *pisiferas* selected for high oil yield following the reciprocal recurrent selection method (Gascon and Berchoux, 1964) or the family and individual selection method (Hardon, 1970). Studies of diversity and the resources involved in breeding programmes have allowed different researchers to propose new strategies for the efficient selection of parental lines (Billotte *et al.*, 2001; Mayes *et al.*, 2000; Purba *et al.*, 2000; Kularatne and Rajanaidu, 2001; Hayati *et al.*, 2004; Norziha *et al.*, 2008; Singh *et al.*, 2008; Cochard *et al.*, 2009). Today, molecular markers are a useful technique to select genetically distant individuals. Microsatellite markers or simple sequence repeats (SSR) offer advantages with respect to other molecular markers. The SSR include co-dominance, multi-alleles and high level of polymorphic information, which, together with their high abundance and random distribution within the genome, make them a valuable tool for studying genetic diversity between closely related cultivars (Kalia *et al.*, 2011). The aim of this study was to molecularly characterize oil palm (*E. guineensis* Jacq.) progenies from different origins using microsatellite markers in order to assess their genetic diversity and provide important information to optimize their use. In this way, families could be chosen to ensure that maximum genetic divergence is incorporated in breeding programmes.

Materials and methods

Plant material

In this study, 193 samples of pre-bred material, represented by 23 families (progenies), were analysed. The resulting palms from controlled pollinated crosses between individuals of different origins are presented in Table 2. The term 'origin' refers to the geographical regions where the *E. guineensis* material was maintained, including the Democratic Republic of the Congo, Ekona, Djongo, Mongana, Yangambi and Brabanta, which were described previously by Corley and Tinker (2003), Corley and Castro (2004) and Castro and Corley (2007).

DNA extraction and amplification of microsatellite markers

Genomic DNA was extracted from young leaves using a Qiagen extraction kit (Ref. 69 106) following the

manufacturer's instructions. DNA quality was evaluated on a 0.8% agarose gel and then quantified by spectrophotometry. The DNA of each sample was diluted to a concentration of 5 ng/μl for use in the amplification reaction of each microsatellite. A total of 20 microsatellite markers were selected and evaluated. According to their location within the genome, these markers are at independent loci and have been mapped to 14 of the 16 linkage groups (Billotte *et al.*, 2001, 2005; Singh *et al.*, 2008). The amplification conditions for each marker were in accordance with those reported by the authors. To visualize the amplification product, a 6% denaturing polyacrylamide gel, 5 M urea and silver nitrate staining were used.

Data analysis

Using the algorithms in the FSTAT software (Goudet, 2002), for each locus, the following genetic diversity parameters were calculated: allelic frequencies; total number of alleles; allelic richness (R_a) (calculated as the total number of alleles present in the population at a given locus). It also included the average polymorphic information content (PIC), a parameter used to measure the informativeness of each genetic marker, which is calculated from allelic frequencies. The following

Table 1. Number of alleles and polymorphic information content (PIC) for the 20 loci

Locus SSR	N	A	R_a	PIC
mEgCIR0008	192	5	4.992	0.601
mEgCIR0009	193	5	5.000	0.610
mEgCIR0018	192	4	4.000	0.355
mEgCIR0046	191	3	2.996	0.230
mEgCIR0067	191	8	7.996	0.710
mEgCIR0219	193	6	6.000	0.689
mEgCIR0230	193	4	4.000	0.614
mEgCIR0254	193	8	7.988	0.712
mEgCIR0465	191	4	4.000	0.518
sEg00066	190	7	7.000	0.732
sEg00067	190	2	2.000	0.433
sEg00125	192	4	4.000	0.334
sEg00126	192	2	2.000	0.481
sEg00127	193	2	2.000	0.095
sEg00140	193	3	3.000	0.347
mEgCIR0802	193	9	9.000	0.744
mEgCIR1730	192	5	5.000	0.676
mEgCIR3282	193	6	5.988	0.756
mEgCIR3363	193	4	4.000	0.705
mEgCIR3546	192	5	5.000	0.593
Total		96	4.798	0.547

SSR: simple sequence repeats; *N*: number of evaluated individuals per locus; *A*: number of detected alleles per locus; R_a: allelic richness.

Table 2. Genetic diversity parameters in *Elaeis guineensis* from different families

Family codes	Origin from the family	N	A	A_e	H_o	H_e	PA
1607	Yangambi × (Deli × Yangambi)	7	2.200	1.609	0.364	0.333	0
1607	Yangambi × (Deli.Yangambi × Djongo)	15	2.350	1.814	0.420	0.406	0
1615	Deli × Yangambi	2	1.900	1.753	0.475	0.467	0
1640	Yangambi × Djongo	24	2.700	1.907	0.503	0.424	0
1911_55	Congo	17	3.450	2.419	0.510	0.522	1
1999	Brabanta × Djongo	24	3.300	2.022	0.428	0.449	2
1694	(Yb × URDeli) × (De.Yb × Dj)	5	2.650	2.159	0.600	0.521	0
1799	(Dj × Yb) × Djongo	4	2.100	1.764	0.438	0.388	0
1869	(Yb × Bra) × (Yb × De.Yb)	3	2.400	2.039	0.600	0.520	0
1871	(Yb × Nf.De) × Mongana	4	2.100	1.790	0.425	0.405	0
1873	Djongo × (De.Yb × Dj)	5	2.150	1.800	0.500	0.416	0
1905	(Yb × De.Yb) × (Yb × Nf.De)	4	2.300	1.896	0.488	0.466	0
1648_51	Djongo	18	2.700	2.001	0.525	0.455	0
1751	Mongana × Congo	7	2.450	1.943	0.543	0.435	0
1598	Djongo × Ekona	4	2.600	2.157	0.600	0.527	1
2713	Ekona × Djongo	15	3.300	2.232	0.463	0.496	1
ECBP20	Ekona × Ekona	8	2.950	2.256	0.418	0.458	0
7092	AVROS × Djongo	3	2.100	1.773	0.483	0.443	0
1667	Mongana	2	1.850	1.660	0.500	0.433	0
2012	Mongana × NIFOR	2	1.900	1.733	0.525	0.458	0
31094_L	Deli *duras* × Djongo	3	2.350	2.016	0.567	0.483	0
31125_E	Deli *duras* × AVROS	12	2.550	2.021	0.550	0.428	1
1167	Deli *duras* × Mongana	5	2.100	1.791	0.560	0.396	0
Total	23	193					

N: number of evaluated individuals per family; A: average number of alleles; A_e: effective number of alleles; H_o: observed heterozygosity; H_e: Nei's (1978) expected heterozygosity; PA: number of private alleles.

parameters were considered to estimate the genetic diversity of each family: (1) the average number of alleles per locus (A); (2) the number of effective alleles per locus (A_e), which is defined as the number of alleles that can be present in a family; (3) the observed heterozygosity (H_o), which is a parameter used to quantify the number of heterozygous genotypes; (4) unbiased expected heterozygosity (H_e), which was obtained according to Nei's (1978) procedure. This information was obtained using the algorithms included in the GenAlEx 6.1 software (Peakall and Smouse, 2006). To determine the genetic differentiation among the families, Nei's (1987) diversity indices and statistical indices F (F_{ST}, F_{IS}) according to Weir and Cockerham (1984), including confidence intervals, were calculated using the FSTAT software (Goudet, 2002). An analysis of molecular variance (AMOVA) was conducted between and within families using Arlequin (Excoffier and Heckel, 2006). The matrix of the genetic distance between families was generated according to the definition by Nei (1972). The dendrogram was generated from the distance genetic matrix and the UPGMA (unweighted pair-group mean arithmetic average) clustering method using the NTSYS pc 2.11L software (Rohlf, 2000). For the statistical values reported per group, 1000 bootstrap replicates were used in the DARwin5 software (Perrier and Jacquemoud-Collet, 2006). A principal coordinate analysis (PCoA) was carried out using the algorithms included in the GENALEX software (Peakall and Smouse, 2006).

Results

Allelic diversity

A total of 96 alleles were obtained in the 193 evaluated samples (Table 1). The number of alleles per locus ranged from 2 to 9. The loci mEgCIR0067, mEgCIR0254, sEg00066, mEgCIR0802 and mEgCIR3282 had the highest number of alleles and thus the highest allelic richness. Also, these loci showed a maximum value of PIC; therefore, these loci are highly informative. The loci sEg00067, sEg00126 and sEg00127 had the lowest number of alleles ($A = 2$), the lowest allelic richness ($R_a = 2.000$) and a minimum value of PIC, suggesting that these loci were least informative.

Genetic diversity of oil palm **E. guineensis** *within each family*

The diversity parameters that were evaluated in the different families are presented in Table 2, where the average number of alleles per locus (A) within

Table 3. Diversity and genetic differentiation coefficients (Nei, 1987) and F statistics (Weir and Cockerham, 1984) among families

	H_o	H_s	D_{ST}	H_T	G_{ST}	G_{IS}	Mean	95% CI
	0.499	0.442	0.115	0.557	0.207***	-0.129		
F_{ST}							0.174***	0.149, 0.198
F_{IS}							-0.091	$-0.171, -0.000$

H_o: average of the observed genetic diversity; H_s: average of the genetic diversity among groups; D_{ST}: average genetic diversity among subgroups; H_T: total genetic diversity; G_{ST}: genetic differentiation coefficient; G_{IS}: endogamy coefficient; CI: confidence interval.
***Values are significantly different ($P < 0.001$).

each family ranged from 1.850 to 3.450. The family from Congo had the maximum number of alleles ($A = 3.450$) and the family from Mongana had the minimum number of alleles ($A = 1.850$). In most of the families, the effective number of alleles (A_e) was close to the observed number of alleles (A), indicating an absence of the rare alleles. On the other hand, the families from Congo, Brabanta × Djongo, Djongo × Ekona, Ekona × Djongo and Deli *dura* × AVROS showed differences between A and A_e. These differences indicate that there are some alleles with low frequency or rare alleles.

When comparing the observed and expected heterozygosity, the observed heterozygosity value was higher in all of the progenies. This was expected due to the fact that the study was conducted on progenies from parents of different geographical regions. The families from Congo, Brabanta × Djongo, Djongo × Ekona, Ekona × Djongo and Deli *dura* × AVROS showed alleles that were found only in each one of these families, and so were defined as private alleles (Table 2). However, it is necessary to note that these parameters were directly related to the size of the sample. Therefore a larger number of samples for each family are needed to confirm these results.

Genetic differentiation between families

The parameters of diversity and genetic differentiation according to Nei (1987) and F statistics (Weir and Cockerham, 1984) among families are given in Table 3.

According to Laurentin (2009), the parameters of genetic differentiation between and within the families may be applied regardless of the number of alleles per locus or the evolutionary forces. The total genetic diversity was $H_T = 0.557$. The average genetic diversity among subgroups (D_{ST}; Table 3) was 0.115 among the families, reflecting greater variability among the families. The genetic differentiation coefficient (G_{ST}; Table 3) was 0.207, and its corresponding F_{ST} statistic was 0.174 with a confidence interval of 0.149 to 0.198, suggesting genetic differentiation among the 23 assessed families. This conclusion was in accord with the results of AMOVA, which showed highly significant differences with a partitioning of 17% genetic diversity (heterozygosity) among the families (Table 4). The fixation index values, G_{IS} and its corresponding F_{IS} statistic were negative, indicating an excess of heterozygotes. The F_{IS} was within a confidence interval of -0.171 to -0.000; however, the values of G_{IS} were not statistically significant.

Genetic relationships

The dendrogram obtained by applying the UPGMA algorithm (Fig. 1) and taking individuals as the unit of analysis allowed to identify seven groups with statistical support based on bootstrap values greater than 50. Deli *dura* × AVROS, Deli *dura* × Mongana and Deli × Djongo origins shared the same group (I). This indicates a genetic closeness among them and an apparent reflection of *Deli* but different from the other African origins. Group II contained cultivars of Mongana and

Table 4. Analysis of molecular variance

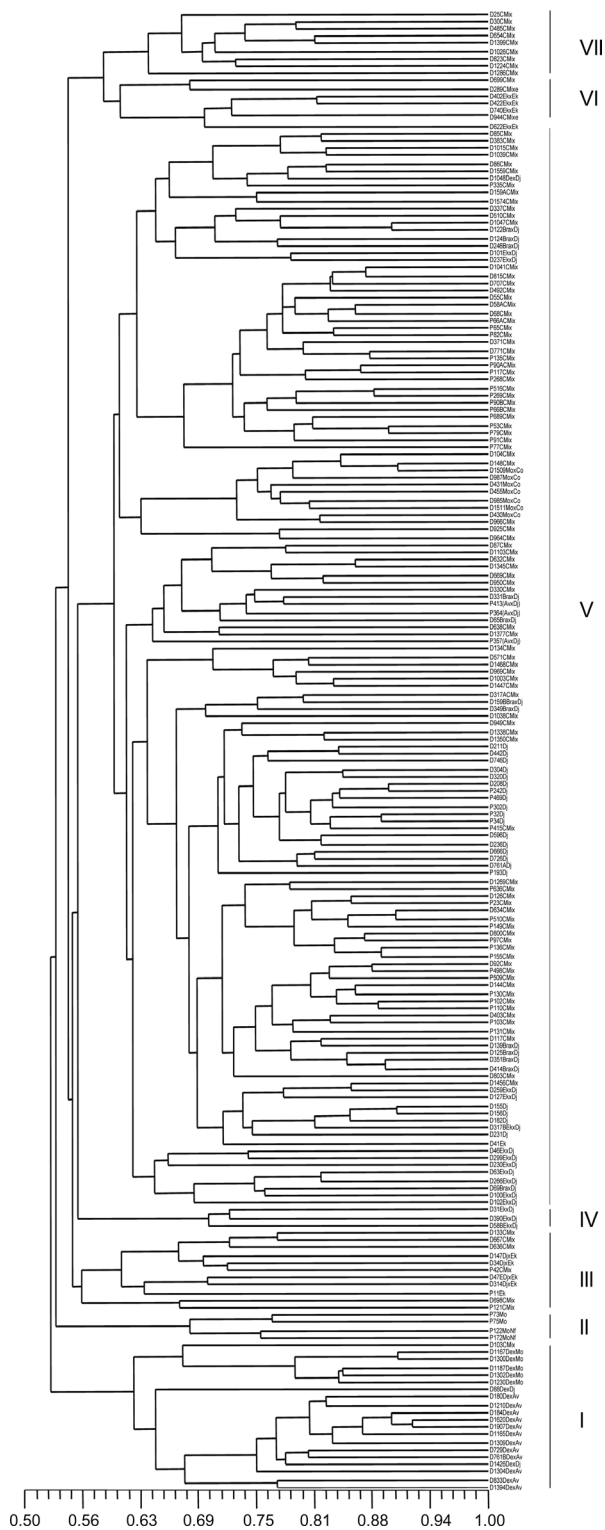| Source of variation	df	Mean squares	Variance components	Percentage of variation
Among families	22	19.41	0.91	17***
Within families	363	4.51	4.51	83
Total	385		5.43	100

*** $P < 0.001$.

Fig. 1. Dendrogram obtained from the Nei and Li (1979) genetic similarity matrix, using the UPGMA (unweighted pair-group mean arithmetic average) grouping method for oil palm materials from different origins.

Mongana × Nifor origin. These were *pisifera*-type palms belonging to 2012 and 1667 codes, which differed from the other groups. Group III was made up of cultivars of Djongo × Ekona origin, one *pisifera* palm of Ekona origin and some palms of Congo Mix origin. Group IV was made up of three palms of Ekona × Djongo origin, code 2713, which clearly differed genetically from the other crossing codes of the same origin in group V. Group V was the largest group, which was made up of closely related cultivars as they shared at least one parent of Djongo origin, as shown in the genealogy reported by Castro and Corley (2007) that included Congo Mix, Djongo, AVROS × Djongo, Brabanta × Djongo, Mongana × Congo and Ekona × Djongo. Group VI was made up of cultivars of Ekona × Ekona origin. Group VII comprised mostly of cultivars of Congo Mix origin, code 1871, with a common parent of NIFOR origin. The Congo Mix origin included different crossing codes. That is why it was distributed in different groups.

In the PCoA, the first two coordinates explained 65% of the total variation (the first one explained 38% and the second one 27%). Four groups were visible in these two coordinates (Fig. 2). Group I was made up of Deli *dura* × AVROS and Deli *dura* × Mongana origins. Group II was made up of origins Ekona × Ekona and Djongo × Ekona. Group III was represented mostly by accessions from Djongo. Group IV comprised mostly of progenies from Mongana.

Discussion

The microsatellite markers used in this study made it possible to distinguish between closely related cultivars, and these markers were generally highly informative with regard to heterozygosity compared with other types of molecular markers used in the oil palm. For instance, for the African origins, the value of expected heterozygosity with the amplified fragment length polymorphism (AFLP) technique was never below 0.65, while with isozymes, the expected heterozygosity values varied between 0.1 and 0.5 (Purba *et al.*, 2000; Hayati *et al.*, 2004). In restriction fragment length polymorphism (RFLP) and AFLP analyses performed in *E. guineensis* accessions, the value of expected heterozygosity was under 0.135 (Barcelos *et al.*, 2002). In the RFLP study, with 359 accessions of oil palm originating from 11 African countries, the value of expected heterozygosity was under 0.25 (Maizura *et al.*, 2006).

In most of the progenies evaluated, the observed heterozygosity value was higher than expected, reflecting a high proportion of heterozygous genotypes within each family. Moreover, the fixation index values, G_{IS} and its corresponding F_{IS} statistic were negative, indicating an excess
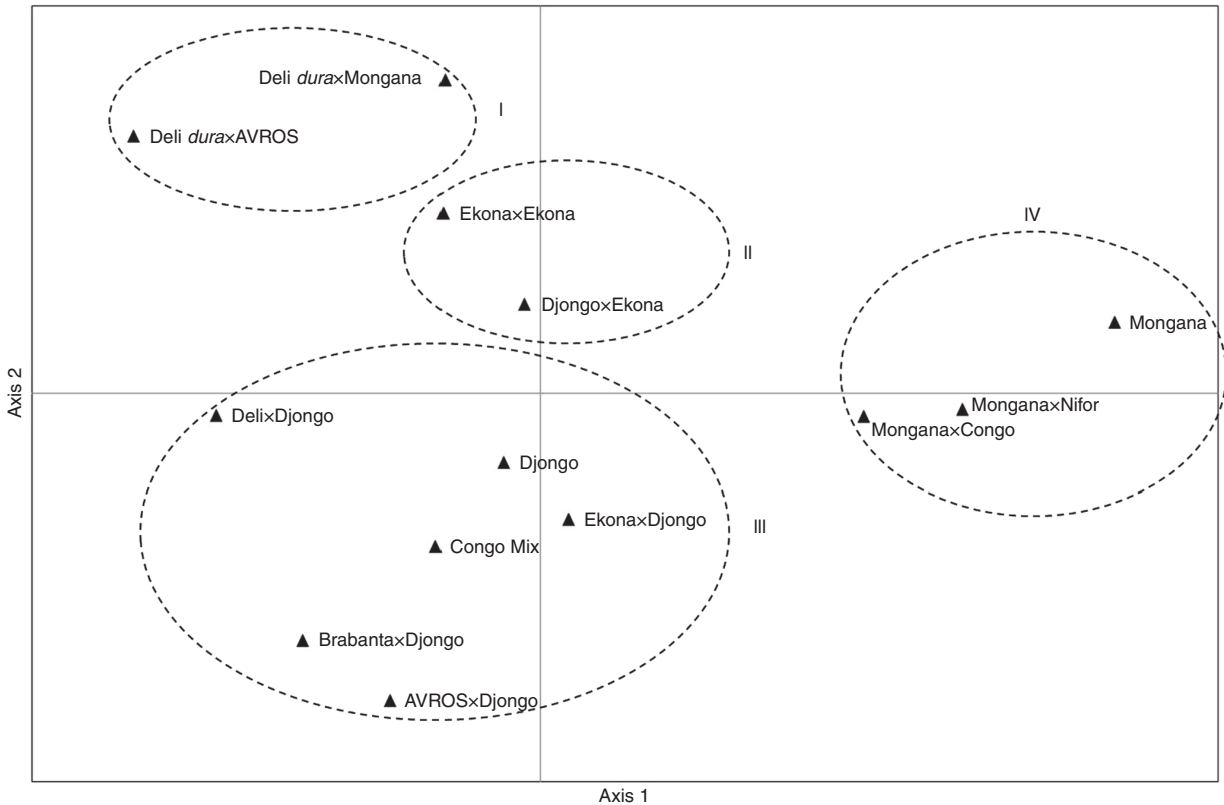
Fig. 2. Principal coordinate analysis based on 20 simple sequence repeats of oil palm *Elaeis guineensis* Jacq. families from different origins.

of heterozygotes in the evaluated progenies. The coefficients ($G_{ST} = 0.207$ and $F_{ST} = 0.174$) and the AMOVA used in this study show that there is genetic differentiation among some of the evaluated families, which was reflected in the groups obtained by the dendrogram and principal coordinate analyses. This difference could be due to some of the families being enriched with germplasm from different origins through controlled pollinated crosses. The PCoA showed more defined groups, which is in agreement with the parent they descended from based on the pedigree used in a pre-breeding programme. The groupings of progenies into one cluster suggest that the families were sampled from the same source of genetic stock. The palms of group I contained material derived from the four palms planted in the Bogor Botanical Gardens in 1848 that derived from Deli *dura* populations, which still form the major basis for the *dura* oil palm germplasm in South-East Asia. According to Rosenquist (1986), the Deli *dura* type is generally considered a breeding population of restricted origin, which is the basis of various oil palm breeding programmes around the world and results from selection through many generations (Corley and Tinker, 2003). In addition, the families of group I were genetically distant with respect to other progenies. This result is consistent

with the results obtained by Cochard *et al.* (2009) who found a strong genetic differentiation ($G_{ST} = 0.231$) between Deli cultivars (Indonesia) versus cultivars of African origin (Zaire). At the phenotypic level, Ooi (1975) found that the cross between Deli *dura* and an unrelated Congo origin increases the additive genetic variance for the number of bunches and average bunch weight, increasing the oil yield potential. The advantages of Deli *dura* × Djongo with 7.5 tons of oil/ha/year and Deli *dura* × Mongana with 6.8 ton/ha/year have been observed in field tests in the Colombian municipality of Cumaral, Meta (Castro and Corley, 2007). This agronomic value has been confirmed under Indonesian conditions, where yields of approximately 7.2 tons of oil/ha/year have been reported for this type of cross (Durand-Gasselin *et al.*, 2000). The parents of group II probably came originally from palms from a village called Lisombe in Cameroon, characterized by having thin-shelled fruits; for this reason, the term *lisombe* became a synonym for *tenera*. One of the first plantations in Cameroon was established in the state of Ikassa around 1919, from seeds of these palms (Corley and Tinker, 2003); these palms exhibited very good bunch composition, an agronomic potential that has been demonstrated under Colombian conditions, where the Ekona × Ekona origin

has recorded yields of 7.3 tons of oil/ha/year. The origins that made up group III came mostly from the famous Djongo palm, which was used to establish one of the first plantations in Palmeraie de la Rive at Yangambi in Zaire in 1922. Beirnaert (1933) described in greater detail the selection work carried out in Congo, which was basically focused on practical and theoretical production issues. An important feature that makes this origin particularly different is the fruit which is elongated and ovoid shaped, with a thin shell, kernel located towards the centre and the mesocarp portion in the basal part of the fruit (Corley and Tinker, 2003). The best lines were established in Binga, a plantation that included a former substation of Yangambi, which led to the development of the Binga breeding programme that also included cultivars of Brabanta and Mongana origin, among others (Hardon *et al*., 1976; Rosenquist 1986 cited by Corley and Tinker, 2003). Thanks to the introgression of other origins, the potential of these cultivars has increased considerably as evidenced by production data from Brabanta × Djongo crosses, with 7.6 tons of oil/ha/year in field tests in the Colombian municipality of Cumaral, Meta (Castro and Corley, 2007).

It is important to highlight that palms of group IV are primarily *pisifera* type from Mongana origins and were the most genetically distant. One of the important objectives of this study was to identify genetic differentiation between the different origins for their use inside breeding programmes. The cost of maintaining oil palm pre-bred material is extremely high. Therefore, the genetic relationships estimated from the molecular data suggest that it could be convenient to select the families more distant from each group and the palms more distant from each family selected to reserve genetic variability in few accessions. Subsequently, the new gene pool can be used into the base population of the first generation of breeding. This information will guide us in the decision-making process when planning breeding programmes focused on crosses to develop new populations with an acceptable broad genetic base and adaptability. In this way, sources of resistance to biotic and abiotic factors can be identified for the development of new varieties with competitive advantages for the sector. Moreover, it would be convenient to identify agronomic traits, such as bunch position in the crown, or yield components such as the percentage of the mesocarp/fruit to obtain new genetic material with a high yield potential.

Acknowledgements

References

Barcelos E, Amblard P, Berthaud J, and 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: 1105–1114.

Beirnaert A (1933) Les base de la sélection du palmier á huile. *Agron Colon* 135–146.

Billotte N, Risterucci AM, Barcelos E, Noyer JL, Amblard P and 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, Herrán A, Asmady H, Billot C, Amblard P, Durand-Gasselin T, Courtois B, Asmono D, Cheah SC, Rohde W, Ritter E and Charrier A (2005) Microsatellite-based high density linkage map in oil palm (*Elaeis guineensis* Jacq.). *Theoretical and Applied Genetics* 110: 754–765.

Castro F and Corley RHV (2007) Advances in the Research and Development program of Unipalma D × P seeds. *PALMAS* 28: 227–240.

Cochard B, Adon B, Rekima S, Billotte N, Desmier de Chenon R, Koutou A, Nouy B, Omoré A, Purba AR, Glazsmann JC and Noyer JL (2009) Geographic and genetic structure of African oil palm diversity suggests new approaches to breeding. *Tree genetics & genomes* 5: 493–504.

Corley RHV and Tinker PB (2003) *The Oil Palm*. 562 pp. Oxford, UK: Blackwell Publishing.

Corley RHV and Castro F (2004) The Unipalma oil palm breeding program. *PALMAS* 25: 311–325.

Durand-Gasselin T, Kouamé R, Cochard B, Adon B and Amblard P (2000) Diffusion vériétale du palmier à huile (*Elaeis guineensis* Jacq.). *Oléagineux Corps Gras Lipides* 7: 207–214.

Dumortier F (2003) Breeding for high yielding progenies at Dami OPRS. In: *Proc. Agric. Conf. Palm oil: The Power-House for the Global Oils & Fats Economy*, pp. 51–74, Malaysian Palm Oil Board, Kuala Lumpur.

Excoffier L and Heckel G (2006) Computer programs for population genetics data analysis: a survival guide. *Nature Reviews Genetics* 7: 745–758.

FAOSTAT (2010) Food and Agriculture Organization of the United Nations – for a world without hunger. Available at http://faostat.fao.org (accessed 19 October 2012).

Gascon J-P and de Berchoux C (1964) Caractéristique de la production d'Elaeis guineensis (Jacq.) de diverses origines et de leurs croisements. Application à la sélection du palmier à huile. *Oléagineux* 19: 75–84.

Goudet J (2002) Institute of Ecology. Biology Building, UNIL software (FSTAT). Version 2.9.3.2. Available at http://www2.unil.ch/popgen/softwares/fstat.htm (accessed 19 April 2012).

Hardon J (1970) Inbreeding in population of the oil palm (*Elaeis guineensis* Jacq.) and its effects on selection. *Oléagineux* 25: 449–456.

Hayati A, Wickneswari R, Maizura I and Rajanaidu N (2004) Genetic diversity of oil palm (*Elaeis guineensis* Jacq.) germplasm collections from Africa: implications from

improvement and conservation of genetic resources. *Theoretical and Applied Genetics* 108: 1274–1284.

Kalia RK, Rai MK, Kalia S, Singh R and Dhawan AK (2011) Microsatellite markers: an overview of the recent progress in plants. *Euphytica* 177: 309–334.

Kularatne RS, Shah FH and Rajanaidu N (2001) The evaluation of genetic diversity of Deli *dura* and African oil palm germplasm collection by AFLP technique. *Trop Agric Res* 13: 1–12.

Laurentin H (2009) Data analysis for molecular characterization of plant genetic resources. *Genetic Resources and Crop Evolution* 56: 277–292.

Maizura I, Rajanaidu N, Zakri A and Cheah S (2006) Assessment of genetic diversity in oil palm (*Elaeis guineensis* Jacq.) using restriction fragment length polymorphism (RFLP). *Genetic Resources and Crop Evolution* 53: 187–195.

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

Nei M (1972) Genetic distance between populations. *American Naturalist* 106: 283–292.

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*. 128 pp. New York: Colombia University Press.

Nei M and Li WH (1979) Mathematical model for studying genetic variation in terms of restriction endonucleases. *Proceedings of the National Academy of Science* 76: 5269–5273.

Norziha A, Rafii M, Maizura I and Ghizan S (2008) Genetic variation among oil palm parent genotypes and their progenies based on microsatellite markers. *Journal of oil palm Research* 20: 533–541.

Ooi SC (1975) Variability in the Deli *dura* breeding population of the oil palm (*Elaeis guineensis* Jacq.). II Within bunch components of yield. *Malaysian Agricultural Research* 50: 20–30.

Peakall R and Smouse P (2006) Genetic analysis in Excel. Population genetic software for teaching and research (GenALEx). Version 6.1. Available at http://biology.anu.edu.au/GenAlEx/Download.html (accessed 19 April 2012).

Perrier X and Jacquemoud-Collet JP (2006) Software (DARwin). Available at http://darwin.cirad.fr./darwin

Purba AR, Noyer JL, Baudouin L, Perrier X, Hamon S and 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. *Theoretical and Applied Genetics* 101: 956–961.

Rohlf FJ (2000) *NTSYSpc: Numerical Taxonomy and Multivariate Analysis System*. Version 2.11L. Setauket, New York: Exeter Software.

Rosenquist E (1986) The genetic base of oil palm breeding populations. In *Proc. Int. Workshop Oil palm germplasm and utilisation*, pp. 27–56, Palm Oil Res Inst Malaysia, Kuala Lumpur.

Singh R, Mohd N, Ting N, Rosli R, Tan S, Leslie E, Ithnin M and Cheah S (2008) Exploiting an oil palm EST database for the development of gene-derived SSR markers and their exploitation for assessment of genetic diversity. *Biologia* 63: 227–235.

Weir BS and Cockerham CC (1984) Estimating F-statistics for the analysis of population structure. *Evolution* 38: 1358–1370.