Int. Symposium on Oil Palm Genetic Resources and Utilization, Z1-Z20

THE GENETIC DIVERSITY OF THE AMERICAN OIL PALM, *ELAEIS OLEIFERA* (KUNTH), CORTÉS REVEALED BY NUCLEAR RFLP MARKERS

Barcelos E.[1], P. Amblard[2], J. Berthaud[3] and M. Seguin[2]

ABSTRACT

*The Restriction Fragment Length Polymorphism (RFLP) technique was used to assess the genetic diversity, its organization and the genetic relationships among 36 American oil palm [*Elaeis oleifera *(Kunth), Cortés] populations from Brazil, Peru, French Guyana, Surinam, north of Colombia and Central America, covering a large part of the natural distribution of the species. Polymorphism in 241 accessions was revealed with 37 nuclear cDNA probes and submitted to Factorial Analysis of Correspondences (FAC), cluster analysis and population genetics analysis (F-statistics). The genetic diversity of* E.oleifera *is strongly structured by geographical origin, with 4 groups clearly distinguished: Brazil, Surinam/French Guyana, north of Colombia/Central America and Peru. Within the Amazon basin, there is a moderate structure that corresponds to the major tributaries of the Amazon river. From the 37 polymorphic RFLP probe/enzyme combinations used, 19 probes (51%) presented simple restriction profiles, with 1 or 2 bands per plant, suggesting a single locus with different alleles, allowing allelic co-dominant coding for them. This led to the identification of 59 alleles for all 19 loci. Three rare alleles, with frequencies lower than 5%, were present in the Brazilian populations, while no other allele was found to be rare or very frequent (>95%) in the whole species, nor in any of the sub-geographic groups. Very low levels of polymorphism and heterozygosity were observed in the non-Brazilian groups: Surinam/French Guyana, north of Colombia/Central America and Peru, with some of these being even monomorphic and completely homozygous for the studied RFLP locus. A high genetic diversity was detected in* E. oleifera *by RFLP nuclear analysis, with a mean of 3.1 alleles in the 19 loci and H Nei = 0.404. These values are higher than the genetic diversity estimated for African* E.guineensis *by using the same probes over 51 accessions covering a broad area of the distribution of this species, and is higher also than in other perennial species studied elsewhere.*

Keywords: *RFLP, genetic diversity, population genetics, tropical perennial species,* Elaeis *oleifera.*

[1] Embrapa – CP 319. CEP 69.011.970 – Manaus/AM. Brazil. Barcelos@cpaa.embrapa.br
[2] CIRAD – CP. BP 3053. – Montpellier CEDEX 134032, France. amblard@cirad.fr / seguin@cirad
[3] IRD/ORSTOM – CP. BP 5045. – Montpellier CEDEX 1 34032. France. berthaud@orstom.fr

INTRODUCTION

The oil palm (*Elaeis guineensis* Jacq.) is an African species that is widely cultivated in most humid tropical regions of Africa, East Asia and the Americas, and represents the second most important source of vegetable oil. Palm oil is widely used in a variety of edible products, with more than 18 million tons being traded worldwide in 1996 (FAO, 1998)

Although oil palm has the highest productivity among the major cultivated oil crops, commercial varieties have a narrow genetic base due to the restricted number of ancestral progenitors (HARTLEY, 1988; SIMMONDS, 1993; HARDON, 195; YUAN, 1996). The major oil palm breeding programs are facing various challenges at present, and broadening the genetic base is a high priority for meeting many of these. New varieties should present disease resistance, low trunk growth and better oil quality.

Oil palm breeders have been interested in the wild species *E. oleifera* for a long time. The first introduction of this species to the African oil palm research centers dates to 1927 (VANDERWEYEN, 1949). More recently, important genetic resources of this species were collected and distributed to the most important oil palm research centers, where they are used in various ways according to the problems faced by local oil palm breeders (LE GUEN, 1991).

Elaeis oleifera is found exclusively in humid tropical America. With *E. guineensis*, it constitutes the genus *Elaeis*. They are cytologically very similar and fertile interspecific hybrids are easily obtained (HARTLEY, 1988). *Elaeis oleifera* presents many interesting agronomic traits, such as resistance to some important diseases and pests, slow trunk height growth and high unsaturated fatty acid content. All these characteristics are desirable in oil palm commercial varieties, so hybridization is used to transfer these traits, since the direct use of this species is limited by its very low oil yield (HARDON, 1969; VALLEJO, 1975; MEUNIER, 1976; AMBLARD, 1995).

Elaeis oleifera is distributed from southern Mexico to the central Brazilian Amazonia where it is found from just west of the mouth of the Tapajós river in the east up to Ecuador in the west (MEUNIER, 1975). Most populations are located near river margins and in flooded areas, in general in more open and sunny habitats. In some regions, such as along the margins of the Madeira River in Brazil, populations are frequently found growing on higher and drier areas with good soil fertility. This type of soil, locally called "terra preta do indio" is very restricted, and considered to be anthropogenic (BARCELOS, 1986). Plants of *E. oleifera* do not found grow in the forest (MEUNIER, 1975).

Genetic resources of *E. oleifera* are still poorly characterized. Little is known about the organization of genetic variability in the species, especially the level of genetic diversity present, and no comparisons exist between populations from geographically distant origins, such as Colombia, Peru, Suriname or French Guyana. In the early 1980s, an important survey of *E. oleifera* genetic resources was carried out in Brazilian Amazon (ANDRADE, 1982; PACHECO, 1982; MEUNIER, 1982). The genetic variability of this material was evaluated using isoenzymes (GHESQUIÈRE, 1987), but little information was obtained at the population level, while the genetic variability were slightly structured according to the major amazon rivers.

The information at population level has important implications for both germplasm utilization and genetic resources conservation. A preliminary study by Jack et al. (JACK, 1995) identified species polymorphism in *Elaeis* using low-copy nuclear RFLP probes but on a very restricted number of accessions and markers. In this study the genetic variability and its structuration at species, region and population levels is intended.

MATERIAL AND METHODS

Plant material and DNA extraction

Two hundred and forty one plants of *E. oleifera*, representing 32 'populations' from Brazilian Amazonia and one 'population' each from Peru, French Guyana, Surinam and Central America (Table 1) were studied. These 'populations' are representative of the natural distribution of the species (Fig.1). Each plant corresponds to a seed collected directly from a wild population.

In this study, a 'population' corresponds to a group of plants growing naturally in the surveyed areas. Two or more collection sites were grouped to constitute a 'population,' in order to obtain a more representative size and a more accurate statistical analysis. This grouping was done after a preliminary genetic analysis of the plants from each collection site if there was no or very low genetic divergence for the loci analyzed and if the collection sites were geographically close. According to this grouping criteria, the Central America 'population' represents plants from northwestern Colombia, Panama, Costa Rica and Nicaragua. The French Guyana population represents two collection sites and the 32 Brazilian populations represent 53 collection sites dispersed throughout Amazonia (Table 1).

Leaf samples were harvested from plants growing at oil palm research stations in Brazil (Embrapa, Manaus) and Côte d'Ivoire (IDEFOR, Abidjan), where large germplasm collections exist. Total DNA was extracted from 3g of fresh or 1g of lyophilized leaves, using the CTAB protocol from Gawel and Jarret (GAWEL, 1991).

RFLP procedures

Total DNA of a sub-set of 6 to 10 plants representing the different geographic origins were digested with 10 restriction enzymes and used to constitute test blots in order to screen RFLP probes for the present study. One hundred ninety homologous and heterologous probes were hybridized onto these test blots (data not shown).

Total DNA of each plant was digested with the selected restriction enzymes. Blots were prepared using 8 µg of total restricted DNA from each plant. To account for possible migration distortions, molecular weight markers of 24.8 kb and 1.5 kb, obtained from *Kpn* I restricted λ phage DNA, were added to each restricted genomic DNA sample just before loading onto the gels. These marker bands were revealed when necessary by adding ^{32}P-αdCTP labeled λ phage DNA to the RFLP probes. Raoul (Appligene) molecular weight marker was also run once per comb.

Restriction fragments were separated by electrophoresis in 0.8% agarose gels (20 x 24 cm) in 1X TBE buffer, for 16 h at 1.5 V.cm^{-1} and then transferred onto nylon membrane (Hybond N^{+} , Amersham), by alkaline blotting (SOUTHERN, 1975). Probes were labeled with ^{32}P-αdCTP using the Amersham Megaprime® commercial kit. Pre-hybridization, hybridization and washes were performed according to Hoisington (HOISINGTON, 1992) . Membranes were then exposed to X-ray film, using two intensifying screens at -80°C for 3 days for homologous probes and up to 10 days for heterologous probes. Each membrane was re-used 10 times or more, after stripping with boiling 1% SDS and agitating at 65°C for 30 minutes on a shaker after each utilization.

RFLP probes

A total of 108 homologous *E. guineensis* cDNA anonymous probes (TREGEAR, 1997) and 82 heterologous nuclear cDNA probes were screened on test blots for their capacity to give clear hybridization signals and clear polymorphic RFLP patterns. The heterologous probes were

from rice (CATO, 1996; KURATA, 1994), oat, maize, barley (GARDINER, 1993.) and coconut (LEBRUN, 1995).

Data analysis

The RFLP fragments or bands were scored in two ways: 1) binary: each band was scored as presence vs. absence for all pairs of probe/enzyme combinations; 2) allelic: bands were scored as alleles for a selected set of probe/enzyme combinations producing simple and clear profiles. Since genetic segregation analysis was not done, the allelic hypothesis was based on the fact that RFLP mapping results showed that 68% of cDNA probes detecting RFLP in *Elaeis sp.* reveals only one locus, which behaves like a diploid species presenting 2n=2x=32 (MAYES, 1997). The banding pattern presented by two interspecific *E. oleifera* x *E. guineensis* F1 hybrid individuals included in this study as controls and their parents also support this assumption. In these cases, alleles present in a given accession were visually identified and received a number ranging from 1 to n, n being the total number of alleles per locus identified over all genotypes. Heterozygous genotypes were recorded when displaying two different levels of bands.

The organization of genetic variability was visualized by performing a multi-variate Factor Analysis of Correspondences (FAC). FAC is based on CHI^2 distance, where markers and genotypes both contribute to the definition of the different factors (BENZECRI, 1973). This analysis was performed on the allelic matrix data, where each allelic variable was scored as: 0 = absence of the allele; 1 = presence in a heterozygous state; 2 = presence in a homozygous state. FAC was also performed on the dominant binary data set after disjunction of the binary matrix data. Application of the FAC algorithm on a disjunctive matrix is also called Multiple Correspondence Analysis (MCA). This analysis was performed using algorithms in the ADDAD software package (LEBEAUX, 1985). With this analysis, robust associations between variables and genotypes are detected, revealing any linkage disequilibrium among these variables and the population. The most informative data are represented on the first two axes, showing the relevant structure of the genetic diversity.

The genetic diversity revealed by RFLP markers was estimated by using the allelic data set at the population level with the classical parameters of population genetics: mean number of alleles per locus, effective number of alleles (Crow and Kimura in (HARTL, 1988), percentage of polymorphic loci, Nei's genetic diversity – HNei (NEI, 1978), and allelic richness (EL MOUSADIK, 1996; HURLBERT, 1971).

The effective number of alleles [$A_e = 1/(1\text{-H Nei})$] estimate was based on the frequency of different alleles detected in a population. Considering that different distributions of allele frequency can result in the same homozygosity, estimate of the effective number of alleles, assuming all alleles equally frequent, allows the comparison of populations presenting a different allelic constitution (HARTL, 1988) .

The allelic richness is defined as the number of different alleles found in a fixed sample size. The estimation of allelic richness allows comparison of the allelic diversity in populations presenting different sample sizes. Allelic richness was computed by using the rarefaction method described by Hurlbert (HURLBERT, 1971) for ecological diversity studies with unequal sample size, using the formula proposed by El Mousadik and Petit (EL MOUSADIK, 1996)

Using the RFLP allelic data, the structure of genetic diversity was evaluated (F statistics) by estimating Wright's fixation index (Fis) based on heterozygote deficiency, which is a measure of the deviation from the Hardy-Weinberg equilibrium and indicates the direction of this deviation. Partitioning of total genetic diversity into within-population and between-population components was also assessed by using Wright's fixation indices. (WRIGHT, 1978). The F-parameters (Fis, Fst, Fit) were computed using Fstat software (GOUDET, 1994). Tests of significance were performed on Fis values using the permutations procedure (ROUSSET, 1995)

proposed in the Fstat software. Divergence among populations or groups of populations was evaluated calculating similarity indices and distances, using allelic co-dominant markers. Dendrograms elaborated with Nei's unbiased genetic distances (NEI, 1978) were also used to visualize possible organization of the genetic diversity among more groups. Distance matrix were calculated using BIOSYS software (SWOFFORD, 1981) and trees were built using ABCD software (CIRAD software).

RESULTS

RFLP polymorphism

From a set of 190 heterologous and homologous cDNA probes screened on test blots, 29 homologous and 8 heterologous probes presenting strong and easy to score polymorphic bands were retained for further analysis. The origin of the selected heterologous probes was rice (5 probes - 4 from Cornell University/USA and 1 from RGP/Japan), maize (2 probes - from Missouri University) and coconut (1 probe from CIRAD CP, France).

The RFLP pattern detected with all 37 selected cDNA probes were first scored as bands. A total of 248 bands were retained for FAC statistical analysis, with no missing values. An average of 6.6 bands was revealed per probe over all individuals. The number of polymorphic bands per geographic group (Brazil, Peru, French Guyana, Surinam, north of Colombia and Central America) is presented in Table 2.

From the 37 polymorphic probe/enzyme combinations used, 19 probes (51%) presented simple restriction profiles, with 1 or 2 bands per plant (Figure 1), suggesting a single locus with different alleles. This lead to the identification of 59 alleles for all loci (Table 3). The mean number of alleles per locus was 3.10 and the mean number of effective alleles was 1.68. Allele distribution per geographic group is presented in Table 3. Three rare alleles, with frequencies lower than 5%, were present in the Brazil group. No other allele was found to be rare or very frequent (>95%) in the whole species, nor in any of the geographic group. Very low levels of polymorphism and heterozygosity are observed in the non-Brazilian groups. French Guyana and Central America groups are even monomorphic and completely homozygous in the allelic co-dominant study, the few polymorphic probes in these groups being unscorable as co-dominant loci.

Genetic variation and structure at the species level

The first two axes of the FAC using the 59 alleles data set represented 46% of the total variation (data not show) and showed a strong geographical structuring, with 4 groups clearly distinguished: Brazil, Surinam/French Guyana, Central America and Peru. The Brazilian group was clearly separated from the others by the first axis of the FAC, which explains 26% of total variation. The FAC using binary bands gave similar results (Figure 2), except that some polymorphism was revealed in Central America and Surinam plants, and the French Guyana and Surinam plants are more differentiated. The structuring revealed by these two FACs are similar and the variability explained by the first axis of both analysis is also equivalent. The grouping revealed by FAC analysis was used to estimate genetic parameters, based on the assumption that they represent natural groups.

Nei's (NEI, 1978) genetic distance estimates (Table 4) between these groups varied from 0.005 to 1.182, with a mean 0.695. The populations from Surinam and French Guyana were undifferentiated, as suggested by the FAC results. Cluster analysis (Figure 3) confirmed that the Brazilian grouping is distant from all other groups, with French Guyana-Surinam, Peru and

Central America being equidistant from each other, with a similar distance of around 0.60. Differences between *E. oleifera* groups are in the same order of magnitude as differences between E. *oleifera* and *E. guineensis* (BARCELOS, 1998).

The Nei's genetic diversity (NEI, 1978) was HNei = 0.404 for *E. oleifera* (Table 3). The genetic variation within *E. oleifera* was analyzed considering the 5 geographical groups: Brazil, Peru, French Guyana, Surinam and Central America. The Brazilian group, the best represented in this analysis with 177 individuals, was the only one presenting considerable diversity (HNei = 0.245), with 2.42 alleles per locus, while the other groups were highly or completely homozygous (Table 3).

The fixation index (WRIGHT, 1978) calculated for *E. oleifera* (Table 3) revealed a strong divergence from Hardy-Weinberg equilibrium (Fis = 0.75), suggesting significant structuring within the species as shown by the FAC . Most of the variability detected was due to the divergence between the 5 groups considered (Fst = 0.736).

Genetic variation at population level in the Brazilian group

Considering the group of Brazilian plants as a whole, a genetic diversity of HNei = 0.245 was detected. The HW disequilibrium (Fis = 0.450) was highly significant, suggesting the existence of some structuring within this group.

A FAC was performed on the Brazilian group with RFLP nuclear allelic co-dominant markers. The first two axes (Figure 4) represented 33% of the total variability and showed some genetic structuring related to the fluvial pattern in Amazonia. The FAC results performed with nuclear RFLP binary data were in accordance with the allelic co-dominant FAC results, showing the same structuring, but with the first two axes representing only 20% of total variability (data not shown). The concordance between these results suggests that the selection of a subset of loci by RFLP multi-allelic scored can be considered a unbiased sample of the overall RFLP diversity.

Overal, the 32 Brazilian populations presented an average of 1.4 alleles per locus and per population, with values ranging from 1.1 to 1.8 allele per locus (Table 5). The percentage of polymorphic loci averaged over all populations was 38%, with the values varying from 11% to 63%, and average Nei genetic diversity was 0.160, ranging from 0.026 to 0.297 (Table 5). The effective number of alleles ranged from 1.03 for the more homozygous populations, up to 1.42 for the more polymorphic ones (Table 5). Allelic richness averaged 0.380, and ranged from 0.079 to 0.693, while a total allelic richness of 1.079 was revealed for *E. oleifera* as a whole (Table 5).

The fixation index (Fis) revealed a significant Hardy-Weinberg disequilibrium in 86% of studied populations, with most populations (64%) presenting a heterozygote deficiency, but 7 out of 32 (22%) presenting a significant excess of heterozygotes (Table 5). Thirty-five percent (Fst = 0.353) of the total variability revealed by nuclear RFLP over 19 loci and 45 alleles was due to the difference among these 32 populations, suggesting some structuring within this group, as also revealed by the FAC results.

The genetic distances between Brazilian populations using nuclear RFLP allelic data calculated and cluster analysis (data not shown) revealed that the genetic diversity within this group is weakly structured, despite an HNei = 0.245. The relatively small genetic differentiation, averaging 0.06 among all populations, indicates that this diversity is due more to intra-population variability (Fst = 0.35). Nevertheless, the relations among these populations are clearly explained by the main rivers in the region (Figure 5).

Indeed, there is a positive relationship between clustering among plants, as shown on FAC (Figures 6), or among populations, as shown on distance analysis (Figure 4), and the localization on the same river. The exceptions were 3 populations (BSO, DEM, MAN) along the Madeira river, which clustered separately from other populations along the same river (Figure 5).

DISCUSSION

Genetic variability of *E. oleifera* was estimated through nuclear RFLP analysis scored as (a) multi-allelic or allelic co-dominant markers and (b) bi-allelic or dominant binary markers. The results were highly concordant in presenting the same genetic diversity structuring, suggesting that the allelic co-dominant coding interpretation used is acceptable despite the lack of segregation studies. By scoring as dominant binary markers, more probes could be utilized and more diversity was revealed within some groups, while multi-allelic scoring had the advantage of allowing estimation of population genetic parameters.

E. oleifera genetic diversity

A high genetic diversity was detected in *E. oleifera* by RFLP nuclear analysis, with a mean of 3.1 alleles in the 19 loci and HNei = 0.404. These values are higher than the genetic diversity estimated for African *E. guineensis* by isoenzymes analysis using 15 loci over 51 accessions covering a broad area of the distribution of this species, with a mean 2.2 alleles per locus and a HNei of 0.262. (GHESQUIÈRE, 1985).

Compared to other tropical species, genetic diversity in *E. oleifera* is much higher than in cocoa (*Theobroma cacao*), another Amazonian perennial with a similar geographical distribution. RFLP analysis of 43 loci in 155 accessions of cocoa revealed only 1.8 alleles per locus and an HNei = 0.28. Like *E. oleifera*, cocoa accessions originating from French Guyana also presented a very low genetic diversity (HNei = 0.065), with only 1.1 alleles per locus revealed by RFLP analysis (LERCETEAU, 1997).

Hevea brasiliensis, another Amazonian species with a more restricted geographical distribution, is genetically slightly more variable than *E. oleifera*, with 3.9 alleles per locus and a genetic diversity of HNei = 0.45 as revealed by isoenzyme analysis of 13 loci in 183 Amazonian accessions (HAMON, 1998) The genetic diversity revealed by RFLP in the same accessions, based on 25 putative loci, was equivalent to that with isoenzymes (BESSE, 1994 SEGUIN. 1996) *Carapa procera*, another tropical forest tree with a similar geographical distribution. presented 1.7 alleles and a genetic diversity of HNei = 0.20 on the 15 loci of 225 accessions analyzed by isoenzymes, over a restricted part of its distribution in French Guyana (DOLIGEZ, 1997).

Isoenzyme studies performed on 38 species of tropical forest trees from different genera revealed in average of 1.9 alleles per locus and a genetic diversity of HNei = 0.19 (HAMRICK. 1992). The genetic diversities presented by *E. oleifera*, *Hevea brasiliensis* and *Carapa procera* are much higher than these average values estimated for other perennial tropical species.

Our results with Brazilian populations of *E. oleifera* presenting the highest genetic variability in the species invalidate the hypothesis of Ghesquiére et al. (GHESQUIÈRE, 1987), who supposed that most of the *E. oleifera* species diversity should be found in Central America or Guyanas.

Genetic structure at the species level

The strong spatial genetic structuring at the species level revealed by the FAC reflects an important genetic differentiation among these groups, as indicated by the Fst value of 74%. The Fst value (Table 6) found for *E. oleifera* is much higher than other values found for population differentiation in tropical perennial species, which vary between 10 and 12% (LOVELESS. 1992). This high value can be attributed mainly to the presence of specific alleles in each group and to the high level of allele fixation in French Guyana, Peru and Central America (Table 3).

For the Brazilian populations, the Fst value calculated from the allelic data set was 35%, indicating variability within this region due to inter-population differentiation. Even within

Brazil, the Fst is still much higher than the values of 10-12% found for tropical perennial species in general (HAMRICK, 1992; LOVELESS, 1992) . Two tropical perennial species, cocoa (LERCETEAU, 1997) and *Hevea brasiliensis* (SEGUIN, 1996), also presented high Fst values (23% and 25%, respectively), in an equivalent geographical area. The level of genetic differentiation between populations presented by these three species, suggesting that Amazonian perennial tree diversity is structured differently than other tropical perennial trees (LOVELESS, 1992).

The very low genetic diversity found in French Guyana, Surinam and Central America populations, with all or nearly all loci being homozygous, was unexpected for an almost strictly allogamous species like *E. oleifera,* suggesting that one or several recent bottleneck events occurred in these areas. The presence of group specific alleles suggests significant genetic drift experienced by populations, followed by an ancient interruption of gene flow among these groups, raising the possibility of an historical discontinuity in the distribution of this species in the American continent.

Genetic structuring at the Brazilian population level

The estimation of Hardy-Weinberg equilibrium deviations at the population level by the fixation index (Fis) indicates that 62% of the Brazilian populations presented an heterozygote deficit (Fis > 0), 22% of the populations had excess heterozygotes (Fis < 0), and only 16% were in Hardy-Weinberg equilibrium. The heterozygote deficiencies may be due to different causes, such as the Wahlund effect, due to artificial grouping of individuals from populations differing in allelic frequency or allele composition; sampling bias; selective environmental factors favoring homozygotes, such as mating among closely related individuals and occasional self pollination events. In the present study, some populations were artificially created in order to get sufficient numbers for genetic analysis, which could account for these Wahlund effect results. The heterozygote excess in adult natural populations is commonly reported in isoenzyme studies (Bush and Smouse, 1992 cited by (DOLIGEZ, 1997) and could also reflect a better fit of the heterozygous individuals. The reduced sample size for some populations could also effect the HW deviation, as there is a greater chance of sampling bias.

Genetic diversity of 124 plants of the same Brazilian origins of *E. oleifera,* as revealed by 14 isoenzyme loci, presented an He = 0.271 and 2.04 alleles per locus (GHESQUIÈRE, 1987), which is similar to the present RFLP allelic results (He = 0.245 and 2.4 alleles per locus). The main difference between the results from the isoenzyme analysis and the present RFLP study is that isoenzyme revealed more genetic diversity in the BR 174/Rio Branco region, while RFLP analysis identified the upper Solimões River region as presenting more genetic diversity.

Among the populations from the Rio Madeira which were collected on anthropogenic soils, three populations (BSO, DEM, MAN) clustered together outside of their geographical group when analysis was performed with RFLP allelic data. This suggests that these populations could have a recent anthropogenic origin via introduction of seeds from other regions, in this case from the Manaus or the Rio Negro regions. When performing analysis with RFLP markers scored as bands, only one population (MAN) of this group clustered outside of its geographical group (Figure 4).

The AGU population, in spite of being geographically close to Manaus, was isolated from all other groups in the analysis performed with allelic co-dominant RFLP markers (data not shown), but clustered in its geographical group when analysis was done with the RFLP binary bands data set. Furthermore, this population was one of the most homozygous among all the Brazilian populations (Table 5) and its genetic divergence could be a consequence of its allelic poverty.

The IGV population is the most western Brazilian population, located in the region of the upper Solimões River near the Peruvian border and is genetically closest to the Peruvian population, suggesting some gene flow between them. It is the most divergent population among all the Brazilian populations.

The Mangenot population (MGT) is the first *E. oleifera* population from Brazil introduced into the ex-IRHO network in the 1960s. It is certainly from Brazil, but is of unknown origin. Our analysis was unable to relate it to one of the Brazilian populations studied.

The genetic differentiation among the Brazilian populations is due mainly to the variation in allelic frequency. Inter-population divergence was found to be more important when populations located along different rivers are compared, but considerable divergence can also be found between more distant populations along the same river (Figure 4), as is the case of the Solimões River. This suggests an important role for rivers in gene flow within *E. oleifera* in Brazil. Unfortunately, no information is available on gene flow and seed dispersal for this species, but considering that *E. oleifera* seed normally cannot float, other dispersion mechanisms must exist, such as transport on other floating objects. The same mechanisms involved in the dispersal of African *E. guineensis*, such as toucans, rodents and, more recently, humans (ZEVEN, 1967; Schnell, 1946, cited by (MALEY, 1998) can be hypothesized to be involved in the American species. Human action may have been especially important, as 1) amerindians used it (?) as an oil source for painting, skin protection, lighting and, probably, food, and 2) because rivers were the main means of transport in Amazonia.

The population differentiation in *E. oleifera* could be attributed to various factors, such as population discontinuity, limited pollen gene flow (restricted by high air moisture and entomophilous pollination), and relatively limited seed dispersal compared to the magnitude of geographic distances in Amazonia.

The RFLP analysis revealed two different levels of genetic differentiation in the *E. oleifera* germplasm studied. The first is at the scale of the continent, where complete reproductive isolation for long periods resulted in very strong group divergence. The continuous distribution of this species from Central America to Peru, through Brazil and the Guyanas, is reported by some authors (MEUNIER, 1975). Nevertheless, large discontinuities in this distribution have been reported between Central and South America (HENDERSON, 1995). Additionally, the Andes and the Guyana plateau constitute geographic barriers that help explain the observed genetic divergence between Central America and the Guyanas. In contrast, there are no geographic barriers between Brazil and Peru, and the genetic differentiation of the Peruvian population is due to unknown ecological or biological factors. Limited seed and pollen dispersion in this region could also result in the divergence between these populations.

One intriguing fact is that the geographical distribution of this species does not reach lower Amazonia, although suitable ecological conditions for its growth are present. The genetic discontinuity between the French Guyana/Surinam and Brazilian populations could be the result of a drier corridor between these two regions during the Pleistocene (SERVANT, 1997), which limited expansion into this geographical area.

The genetic diversity of *Elaeis oleifera* and its structuring revealed in this study will certainly enhance the use and conservation of its genetic resources, as well as help plan new germplasm collections for this species in Amazonia.

Bibliography

AMBLARD, P.; J. M. NOIRET; B. KOUAMÉ; F. POTIER and B. ADON. (1995). Performances comparées des hybrides interespécifiques et du matériel commercial *E. guineensis*. *OCL* **2**, 335 - 340.

ANDRADE, E. B. (1982). "Expedição para coleta de germoplasma de caiaué (Elaeis oleifera (H.B.K.), Cortés) na Amazonia brasileira.,". Embrapa, Manaus.

BARCELOS, E. (1986). Características genético-ecológicas de populações naturais de caiaué (Elaeis oleifera (H.B.K.), Cortés) na Amazônia brasileira. M.Sc., Fundação Universidade do Amazonas, Manaus-AM.

BARCELOS, E. (1998). Étude de la diversité génétique du genre *Elaeis* (*E. oleifera* (Kunth) Cortés et *E. guineensis* Jacq.) par marqueurs moléculaires (RFLP et AFLP). Doctorat, Université de Montpellier II, Montpellier, France.

BENZECRI, J. P. (1973). "L'analyse des données. 2. L'analyse des correspondances.," Dunod, Paris.

BESSE, P.; SEGUIN, M.; LEBRUN, P.; CHEVALIER, M.H.; NICOLAS, D. et al. (1994). Genetic diversity among wild and cultivated populations of *Hevea brasiliensis* assessed by nuclear RFLP analysis. *Theoretical and Applied Genetics* **v.88**, p.199-207.

CATO, S. A. and T. E. RICHARDSON. (1996). Inter and intraspecific polymorphism at chloropast SSR loci and the inheritance of plastids in *Pinus radiata*. *Theoretical and Applied Genetics* **v.93**, p.587-592.

DOLIGEZ, A. and H. I. JOLY. (1997). Genetic diversity and spatial structure within a natural stand of a tropical forest tree species, *Carapa procera* (Meliaceae), in French Guiana. *Heredity* **v.84**, p.72-82.

EL MOUSADIK, A. and R. L. PETIT (1996). High level of genetic differentiation for allelic richness among populations of the argan tree *Argania spinosa* (L.) skeels endemic to Morocco. *Theoretical and Applied Genetics* **v.92**, p.832-839.

FAO (1998). Palm Oil World Production. , Vol. 1999. FAO.

GARDINER, J. M.; e. h. COE; S. MELIA-HANCOCK; D. A. HOISINGTON and S. CHAO. (1993.). Development of a core RFLP map in maize using an immortalized F2 population. ***Genetics* v.134**, 917-930.

GAWEL, N. J. and. R. L. JARRET. (1991). A modified CTAB DNA extraction procedure for *Musa* and *Ipomea*. *Plant Molcular Biology Reporter* **9**, 262 - 266.

GHESQUIÈRE, M. (1985). Enzyme polymorphisme in oil palm (*Elaeis guineensis* Jacq.). II. Variability and genetic structure of seven origins of oil palm. ***Oléagineux* v.40**. p.529-540.

GHESQUIÈRE, M.; E. BARCELOS; M. M. SANTOS; P. AMBLARD. (1987). polymorphisme enzymatique chez *Elaeis guineensis oleifera* H.B.K. Cortés (*Elais melanococca*). Analyse des populations du bassin amazonien. ***Oléagineux* v.42**, p.143-153.

GOUDET, J. (1994). FSTAT, a program for IBM PC compatibles to calculate Weir and Cockerham's (1984) estimators of F-Statistics. .

HAMON, S. DUSSERT; J. CHANTEREAU, J.; M. H. CHEVALLIER; M. DEU et al. (1969). Effects of quantitative and qualitative Principal Component Score Strategies (PCSS) on the structure of coffee, rice, rubber tree and sorghum core collections. ***Genetics Selection Evolution*** *(sous press)*.

HAMRICK, J. L.; M. J. W GODT and S. SHERMAN-BROYLES. (1992). Factors influencing levels of genetic diversity in woody plant species. ***New Forests* v.6**, p.95-124.

HARDON, J. J. (195). Oil Palm. *In* "Evolution of crop plants" (J. S. a. N. W. Simmonds, ed.), pp. 395 - 399. Longman Scientific & Technical, Harlow, Englad.

HARDON, J. J. (1969). Interspecific hybrids in the genus *Elaeis*. II. Vegetative growth and yield of F1 hybrids of *Elaeis guineensis* x *E. oleifera*. ***Euphytica* v.18**, p.380-388.

HARTL, D. L. and A. G. CLARK. (1988). "**Principles of populationsgnetics.**,", Massachusetts.

HARTLEY, C. W. S. (1988). "**The oil palm.**," 3th./Ed. Longman., Essex.

HENDERSON, A. G. GALEANO and R. BERNAL. (1995). "**Field guide to the palms of the Americas.**," Princeton University Press., Princeton.

HOISINGTON, D. (1992). "**Laboratory protocols.**," CIMMYT., México.

HURLBERT, S. J. (1971). The nonconcept os species diversity: a critique and alternative parameters. ***Ecology*** **v,52**, p.577-586.

JACK, P. L.; T. A. F. DIMITRIJEVIC and S. MAYES. (1995). Assessment of nuclear, mitochondrial and chloroplast RFLP markers in oil palm (*Elaeis guineensis* Jacq.). ***Theoretical and Applied Genetics*** **v.90**, n.643-649.

KURATA, N.; G. MOORE; Y. NAGAMURA; T. FOOTE; M. YANO et al. (1994). Conservation of genome structure between rice and wheat. ***Bio/Technology*** **v.12**, p.276-278.

LE GUEN, V. P. AMBLARD; A. OMORE; A. KOUTOU and J. MEUNIER. (1991). Le programme hybride interespécifique *Elaeis oleiferax elaeis guineensis* de l'IRHO. ***Oléagineux*** **v.46**, p.479-487.

LEBEAUX, M. O. (1985). ADDAD. ADDAD, Paris.

LEBRUN, P.; L. BAUDOUIN; M. SEGUIN; Y. P. N'CHO and R. BOURDEIX. (1995). Study on the genetic diversity of coconut palm by means of RFLP. ***OCL*** **v.2**, p.418-421.

LERCETEAU, E. T. ROBERT; V. PETIARD and D. CROUZILLAT. (1997). Evaluation of the extent of genetic variability among *Theobroma cacao* accessions using RAPD adn RFLP markers. ***Theoretical and Applied Genetics*** **v.95**, p.10-19.

LOVELESS, M. D. (1992). Isozyme variation in tropical trees: pattern of genetic organization. ***New Forests*** **v.6**, p.67-94.

MALEY, J. and P. BRENAC. (1998). "**Les variations de la vegetation et des paleoenvironnements du sud Cameroun au cours des derniers millenaires.** Etude de l'expansion du palmier à huile.," Université Cameroun., Yaoundé.

MAYES, S.; P. L. JACK; D. F. MARSHALL and R. H. V. CORLEY. (1997). Construction of a RFLP genetic linkage map for oil palm (*Elaeis guineensis* Jacq.). ***Genome*** **v.409**, p.116-122.

MEUNIER, J. (1975). Le palmier à huile américain *Elaeis melanococca*. ***Oléagineux*** **v.30**, p.51-61.

MEUNIER, J. (1976). Prospections of the Palmae. A necessity for the improvement of oil-yielding palms. ***Oléagineux*** **v.31**, p.153-157.

MEUNIER, J. and P. AMBLARD. (1982). "Prospection *E. melanococca* Brésil,". IRHO, Montpellier.

NEI, M. (1973). Analysis of gene diversity in subdivided populations. ***Proceedings of the National Academy of Science, Washington, DC*** **v.70**, p.3321-3323.

NEI, M. (1978). Estimation of average heterozygosity and genetic distance from a small number of individuals. ***Genetics*** **v.89**, p.583-590.

PACHECO, A. R. (1982). "Expedição para coleta de germoplasma de caiaué (*Elaeis oleifera* (H.B.K.) Cortés), no Estado do Amazonas.,". Embrapa, Manaus.

ROUSSET, F. and. M. RAYMOND. (1995). Testing heterozygote excess and deficiency. ***Genetics*** **140**, 1413 - 1419.

SEGUIN, M.; P. BESSE; D. LESPINASSE; P. LEBRUN; M. RODIER-GOUD et al. (1996). *Hevea* molecular genetics. ***Plantations, Recherche, Développement*** **3**, p.77-88.

SERVANT, M.; M. L. ABSY; A. CARNEIRO FILHO; A. CHAUVEL; R. CORDEIRO et al. (1997). L'evolution à long terme de la forêt Aamazonienne - biodiversité et changements globaux. *In* "**Envrionnement et développement en Amazonie brésilienne.**" (H. THÉRY, ed., ed.), pp. p.166-179. Editons Belin., Paris.

SIMMONDS, N. W. (1993). Introgression and incorporation. Strategies for the use of crop genetic resources. ***Biological Review*** **68**, p.539-562.

SOUTHERN, E. M. (1975). Detection of specific sequences among DNA fragments separated by gel electrophoresis. ***Journal of Molecular Biology*** **v.98**, p.503-517.

SWOFFORD, D. L. and R. B. SELANDER. (1981). BIOSYS-1: a FORTAN program for the comprehensive analysis of electrophoretic data in population genetics and systematics. ***Journal of Heredity*** **v.72**, p.281-283.

TREGEAR, J. (1997). Personal communication.

VALLEJO, G. and D. C. CASSALET. (1975). Perspectivas del cultivo de los hybridos interespecificos de noli (*Elaeis oleifera* (H. B. K.) Cortez) x palma africana de aceite (*Elaeis guineensis* Jacq.) en Colombia. ***Revista ICA*** **10.**

VANDERWEYEN, R. and O. ROELS. (1949). Les varietés d'*Elaeis guineensis* Jacquin du type albescens et l'*Elaeis melanococca* Gaertner (em. Bailey). *Publications d l'I.N.E.A.C. Série scientifique.*

WRIGHT, S. (1978). "**Variability within and among natural populations.**," University of Chicago Press., Chicago.

YUAN, Y. Y. and C. K. WENG. (1996). Guthrie Chemara planting materials. *In* "**Sourcing of oil palm planting materials for local and overseas joint-ventures.**" (A.-H. ISOPB.. ed.), ISOPB., Selangor.

ZEVEN, A. C. (1967). "**The semi-wild oil palm and its industry in Africa.**," State Agriculture University. Institute of Plant Breeding., Wageningen.

Table 1 - Accessions of *E. oleifera* used in the RFLP genetic diversity studies

Groups	Country of origin	Points of collect	Population[1]	Number of accessions
Brazil	Brazil	59	32	177
Central America	Colombia	5		16
	Costa Rica	3	1	9
	Nicaragua	1		4
	Panama	8		14
French Guyana	French Guyana	2	1	10
Surinam	Surinam	2	1	6
Peru	Peru	1	1	5
Total		81	36	241

[1] see "Material and Methods for details

Table 2 : Gene diversity (Nei, 1973) revealed by 248 nuclear RFLP bands on 241 *E. oleifera* accessions.

Group	Sample Size	Number of polymorphic fragments Total	Specific	Diversity*
E. oleifera	241	248		0,225
Brazil	177	186	33	0,157
French Guyana	10	36	4	0,035
Central America	43	35	6	0,040
Peru	5	27	8	0,037
Surinam	6	24	3	0,029

* Nei diversity index at intra group level (Nei, 1973)

Table 3 : Genetic variability by groups of *E. oleifera* revealed by nuclear RFLP co-dominant markers over 19 loci.

Group	Number of Accessions	Polymorphic locus - %	Number of Alleles Per group	Per locus	Effective	Fixed	Specific	Nei Diversity*	Fis
E.oleifera	241	100	59	3,10	1,68	0		0,404	0,75**
Brazil	177	74	46	2,42	1,32	5	24	0,245	0,45**
French Guyana	10	0	19	1,00	1,00	19	5	0,000	-
Central America	43	0	19	1,00	1,00	19	4	0,000	-
Peru	5	10	20	1,10	1,03	18	3	0,029	-0,09**
Surinam	6	5	20	1,10	1,03	18	5	0,026	0,33**

*Nei, 1978 ** significatif at 99% level (Prob Fis/Fst/Fit ho >0.0002) by allele permutation within samples

Table 4 : Genetic divergence among *E. oleifera* groups revealed by Nei's unbiased genetic distance and Fst based on nuclear RFLP alleles

Group	Brazil	French Guyana	Central America	Peru	Surinam
Brazil	*****	0,721	0,718	0,625	0,704
French Guyana	1,182	*****	1,000	0,980	0,344
Central America	0,854	0,642	*****	0,994	0,993
Peru	0,654	0,637	0,617	*****	0,944
Surinam	1,113	0,005	0,596	0,651	*****

Nei's Unbiased Genetic Distance - Below diagonal Fst - Above diagonal

Table 5 - Genetic diversity of *Elais oleifera* revealed by nuclear RFLP analyses as co-dominant and dominant markers,

Pop	Sample size	Co-dominant markers						Dominant markers	
		polymorphic locus - %	Alleles per locus	effective	Allelic Richness	HNei's Diversity*	Fis	Polymorphi locus - %	Nei's diversi index**
1 ITA	10	42	1.5	1.18	0.352	0.149	0.121***	35	0.116
2 CAL	9	37	1.5	1.21	0.441	0.171	0.671***	33	0.112
3 AGU	4	11	1.1	1.03	0.079	0.026	0.000 ns	15	0.056
4 MAN	5	58	1.7	1.39	0.636	0.278	0.500***	36	0.131
5 LIB	5	63	1.8	1.42	0.693	0.297	0.462***	37	0.134
6 DEM	3	47	1.7	1.40	0.684	0.284	0.437***	23	0.093
7 IGA	4	37	1.4	1.24	0.402	0.192	0.200***	22	0.087
8 MAT	4	37	1.4	1.19	0.378	0.160	-0.083***	17	0.063
9 SHE	5	47	1.5	1.23	0.428	0.189	0.062 ns	30	0.116
10 ITP	3	37	1.4	1.22	0.368	0.179	-0.368***	16	0.064
11 BAC	4	37	1.4	1.20	0.359	0.169	-0.105	16	0.069
12 SEV	6	32	1.4	1.16	0.326	0.140	0.136***	27	0.105
13 BAL	6	42	1.5	1.18	0.395	0.155	0.101***	29	0.108
14 NES	5	16	1.2	1.08	0.167	0.074	0.158***	23	0.089
15 SSE	4	32	1.4	1.17	0.336	0.148	-0.182***	17	0.063
16 ALA	4	16	1.2	1.04	0.118	0.039	0.000 ns	17	0.059
17 AMA	4	42	1.5	1.26	0.453	0.205	0.258***	25	0.099
18 CRI	7	37	1.4	1.12	0.274	0.104	0.143***	21	0.067
19 ENS	6	42	1.6	1.23	0.454	0.187	-0.137***	25	0.095
20 BSO	4	47	1.5	1.30	0.494	0.233	0.297***	24	0.095
21 MAU	5	32	1.3	1.15	0.278	0.129	0.020 ns	25	0.089
22 BRA	3	16	1.2	1.08	0.158	0.077	0.600***	14	0.057
23 BRB	4	32	1.3	1.16	0.235	0.135	0.032 ns	19	0.078
24 BRC	13	32	1.4	1.10	0.189	0.095	0.326***	27	0.086
25 BCS	7	47	1.5	1.24	0.444	0.194	0.478***	32	0.117
26 AVE	6	37	1.4	1.19	0.371	0.157	-0.193***	25	0.098
27 IAC	5	42	1.5	1.21	0.417	0.171	0.084***	23	0.086
28 ACA	2.9	42	1.4	1.28	0.421	0.221	0.154***	22	0.092
29 IGV	5	58	1.7	1.33	0.607	0.247	0.116***	37	0.135
30 TEF	4.9	47	1.5	1.21	0.386	0.172	0.082***	35	0.136
31 LMI	5	26	1.4	1.12	0.288	0.110	-0.171***	27	0.095
32 MGT	3	37	1.4	1.20	0.421	0.168	0.077***	16	0.059
Brazil	177	74	2.4	1.32	0.633	0.245	0.450***	74	0.184
E. oleifera	241	100	3.1	1.68	1.079	0.404	0.750***	100	0.267

* HNei (Nei, 1978)

** Nei diversity index (Nei, 1973)

** *significatif at 99% (Prob Fis/Fst/Fist no >0=0.0002), by allele permutation within samples

ns - no significatif

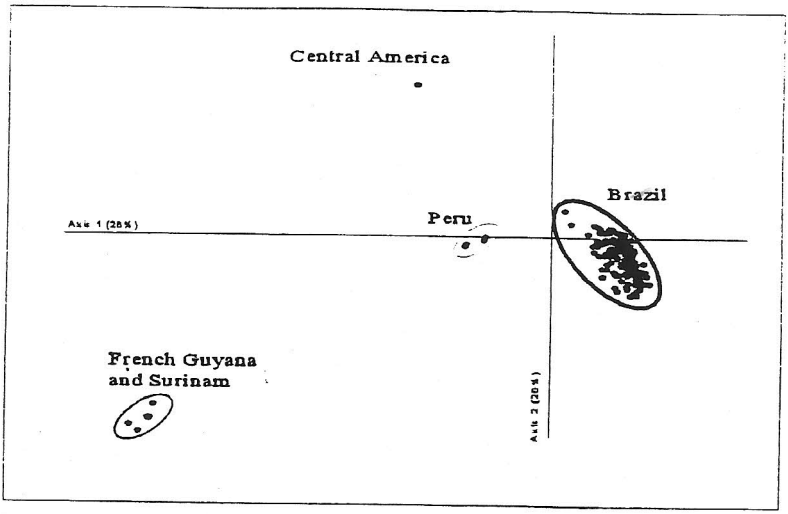


Figure 2 : FAC performed on 59 RFLP alleles revealed in 19 loci on 241 *Elaeis oleifera* accessions.
First plane (Axis 1-2) representing 46% of overall variability.

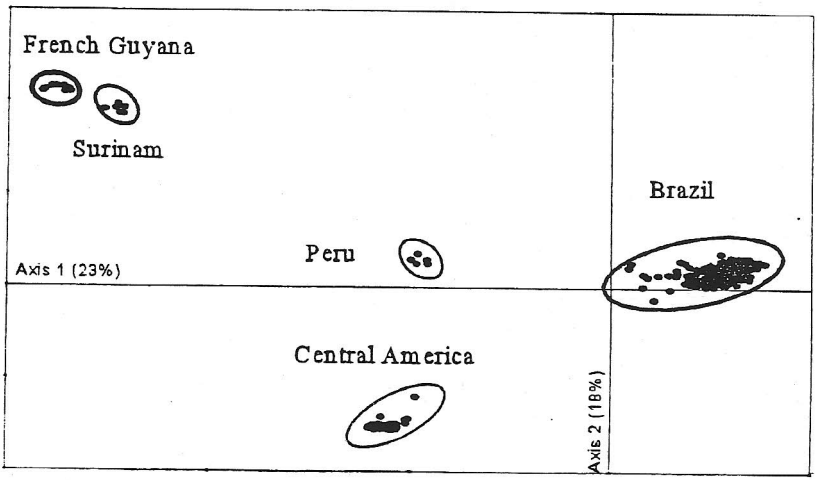
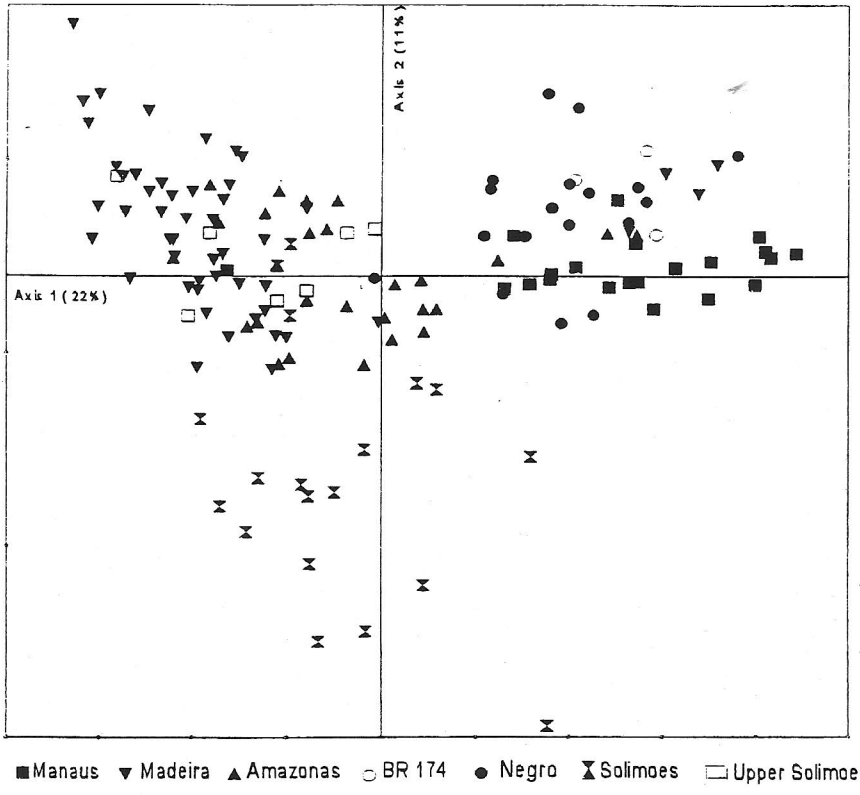
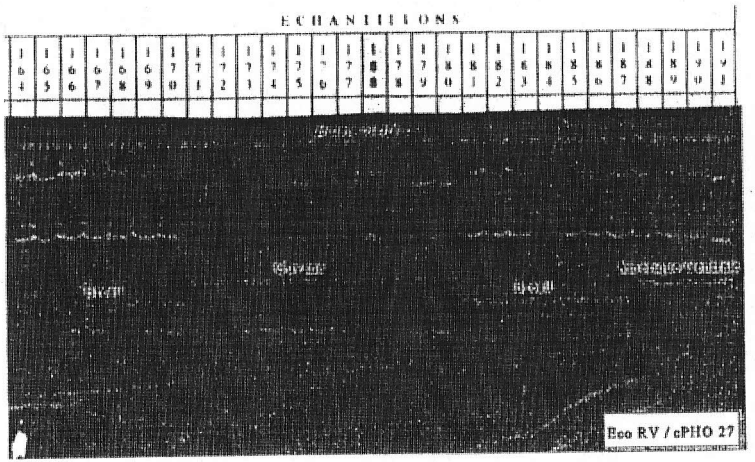
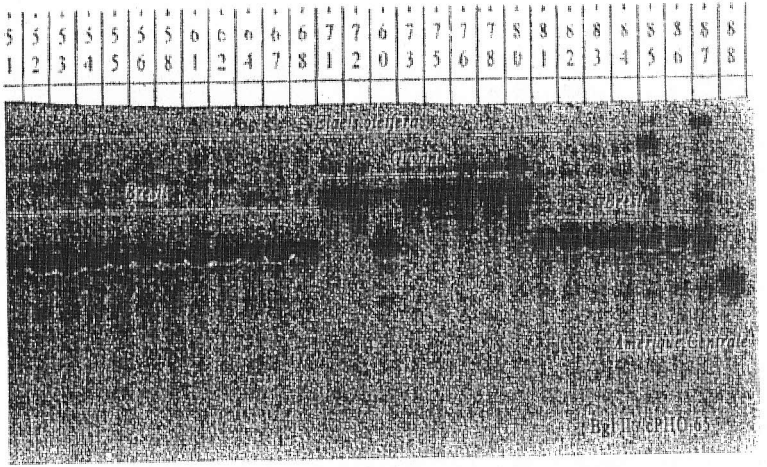


Figure 3 : FAC performed on 248 nuclear RFLP dominant markers revealed by 37 cDNA probes on 241 accessions of *Elaeis oleifera*.
First plane (Axis 1-2) representing 41% of total variability



igure 4 : FAC performed with 45 RFLP alleles over 19 loci on 153 accessions of *E. oleifera* from Brazil. Plane 1-2 representing 33% of overall variability.
* non active individuals





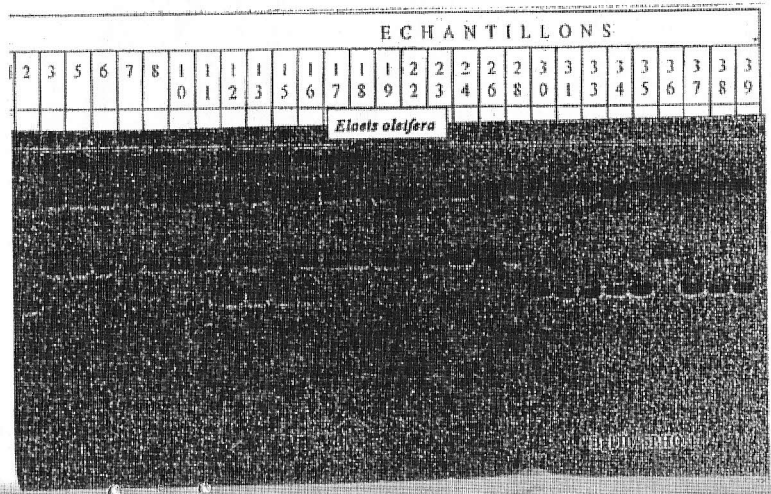


Figure 1 : Autoradiographies après hybridation avec des sondes homologues (*E. guineensis*) présentant du polymorphisme RFLP chez l'*E. oleifera*.

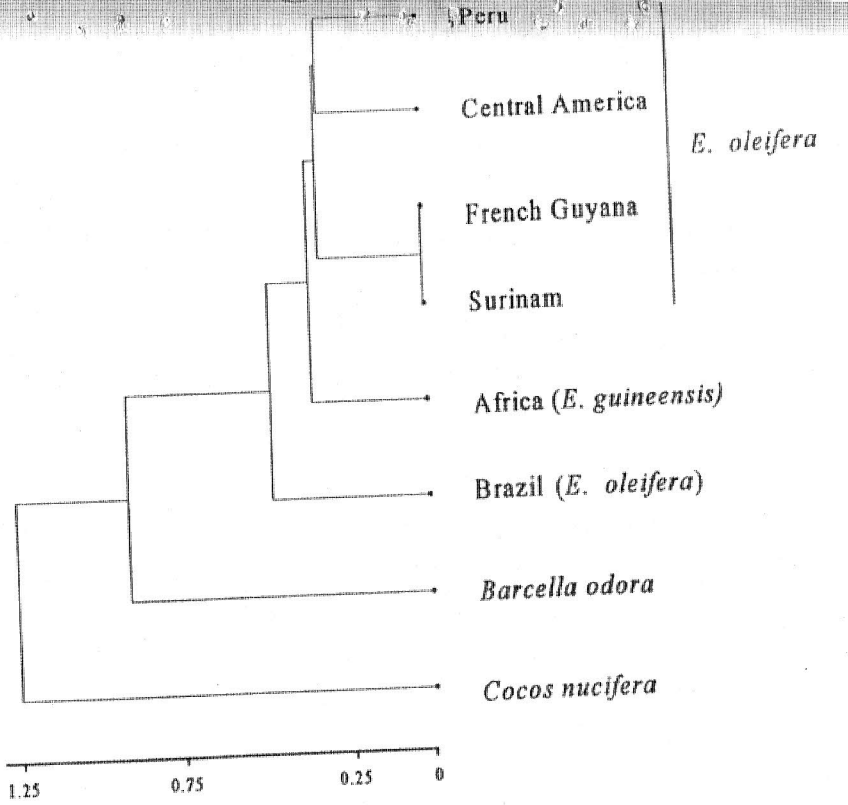


Figure 6 : Genetic relationships between *Cocos nucifera*, *Barcella odora*, *Elaeis guineensis* and 5 groups of *E. oleifera* species. Dendrogram based on Nei (1978) genetic distance calculated on RFLP allele frequencies per group or species.

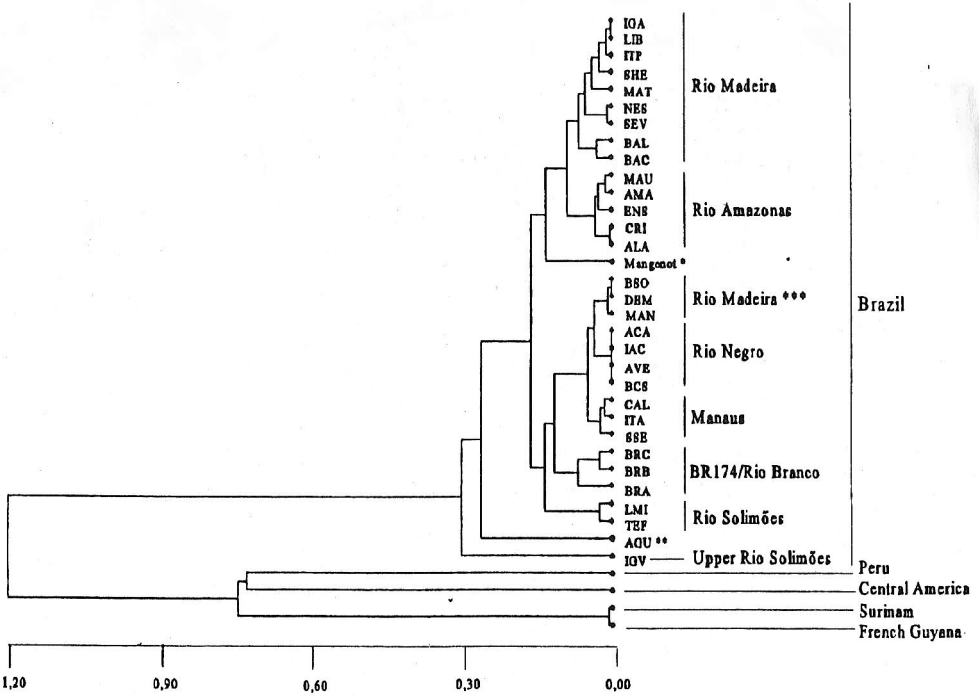


Figure 5 : Relationships among *E. oleifera* populations. UPGMA tree based on Nei's unbiased genetic distances (Nei, 978) calculated with 59 RFLP allel revealed on 19 loci in 36 populations.

* Mangenot, population from Brazil with uncertain origin; ** Population near Manaus ; *** Populations based on "Terra preta do indio."