Theor Appl Genet (1995) 90:561–570

L. Baudouin · T. V. Cao · A. Gallais

Analysis of the genetic effects for several traits in oil palm (*Elaeis guineensis* Jacq.) populations. I. Population means

Received: 17 June 1994 / Accepted: 8 September 1994

Abstract An oil palm experiment was set up in the Ivory Coast to compare the effects of crossing and selfing within two origins, *Deli* and *La Mé*, on the mean and the variability of *Deli* × *Mé* between-origin hybrids. The originality of the experiment lay in the crossing plan, which provided access to genetic parameters related to additivity, dominance and different components of epistasis. This first part covers the analysis of the components of the mean. The parents used were obtained from four palms, two from each origin. Those of *La Mé* origin were half-sibs. The common parent came from a wild stand in the Ivory Coast. Those of *Deli* origin were from two different populations bred in Southeast Asia for several generations from a narrow genetic base. These four parents gave rise to nine *Deli* × *La Mé* hybrid populations with double-cousin-type links. The additive component is more important within the *Deli* origin than within the *La Mé* origin. This may be explained by the large genetic divergence between the two *Deli* parent palms. On the other hand, the additive*additive epistasis is more substantial within the *La Mé* origin, probably because of inbreeding. The discussion concentrates on how this information should be used when choosing parents to be crossed and tested and to produce improved populations. The crossing plan proposed can be of general use and is suitable for other species in a reciprocal recurrent selection programme.

Key words Population structure · Genetic components of the mean · Crossing plans · Reciprocal recurrent selection · *Elaeis guineensis* Jacquin

Communicated by G. Wenzel

L. Baudouin · T. V. Cao(✉)
CIRAD-CP, B.P. 5035, 34032 Montpellier Cedex 1, France

A. Gallais
CNRS-INRA-UPS, Station de Génétique Végétale, Ferme du Moulon, 91190 Gif-s/Yvette, France

Introduction

A reciprocal recurrent selection (RRS) method involving full-sib crosses has been developed by the IRHO[1] for oil palm (*Elaeis guineensis* Jacq.) that will enable the breeder to make the best possible use of sources of variation and lay the foundations for varietal creation. The method consists in jointly improving two complementary groups, with each one used as a tester for the other (Meunier and Gascon 1972). Group "A" provides large bunches and currently includes two origins (*Deli* and *Angola*). Group "B" is characterized by a large number of small bunches and includes numerous origins from West and Central Africa (Gascon and de Berchoux 1963). Varietal creation consists in reproducing the best families from selfed progenies of the parents (S1) or their progenies in within-group crosses (S0) (Jacquemard et al. 1981). Vegetative propagation has opened the way for clone creation and should enable further progress relative to that possible with seeds insofar as it would only reproduce the best individuals from the best crosses. In this way, the technique would exploit all available genetic variability (additivity, dominance and epistasis) and the favourable effects of linkage disequilibria. A specific crossing plan has been drawn up between the selves and cross of two palms from one origin and those from the other origin. These four types of hybrids (S0 × S0, S0 × S1, S1 × S0 and S1 × S1) make it possible to study and compare the relative significance of non-additive effects (dominance and epistasis) and additive effects. This study, which covers a narrow genetic base but is nevertheless representative of the material being bred, could in retrospect justify the choice of a breeding and varietal creation scheme. The present article only gives the results in terms of the mean for the different types of hybrid crosses produced, while the results in terms of variance will be covered in a subsequent article.

[1] Institut de Recherche pour les Huiles et Oléagineux (now CIRAD-CP)

Materials and methods

Planting material

There are three types of oil palm fruit. The *dura* type is characterized by nuts with a thick shell, the *pisifera* type has no shell but is affected to varying degrees by female sterility, and the *tenera* type, which is the most interesting, since it is normally fertile, and has a thin shell, hence a high proportion of mesocarp and kernel (oil-bearing parts of the fruit). This trait is monogenic (Beirnaert and Vanderweyen 1941). Both *dura* and *pisifera* types are homozygous (*Sh*+ *Sh*+ and *Sh*− *Sh*− respectively), while the *tenera* type is heterozygous. The *pisifera* type is of no agronomic value but is used to pollinate *dura* mother palms to obtain *tenera* progenies. For the crossing plan described here, two *dura* palms of *Deli* origin and two *tenera* palms of *La Mé* origin were chosen (Fig. 1).

Crossing plan

Four S1 progenies were obtained by selfing the four palms and two S0 progenies by within-origin crossing. Twenty-four parent palms were chosen from each *Deli* progeny and *12* from each *La Mé* progeny. All of the *Deli* parents had the *dura* genotype, but the *La Mé* parents segregated for fruit type. The *La Mé dura* palms were systematically eliminated from the crossing plan as they did not bear the *Sh* − gene. These six progenies gave rise to nine between-origin factorial combinations (Table 1). Each progeny was involved in three different combinations, and the same individuals were used in each. A combination included 24 full-sib crosses carried out according to a hierarchical plan in which each *La Mé* parent was crossed with 2 *Deli* parents. Each *La Mé* parent was therefore crossed with 6 different testers, while each *Deli* parent was crossed with only 3 testers. The crosses from a given combination is a population of double cousins.

Table 1 Notation of the nine hybrid populations tested. The first index stands for the origin of the *Deli* parents, while the second index stands for the origin of the *La Mé*

		La Mé parental progenies		
		BS1	B × B′	B′S1
Deli	AS1	P_{00}	P_{01}	P_{02}
Parental	A × A′	P_{10}	P_{11}	P_{12}
Progenies	A′S1	P_{20}	P_{21}	P_{22}

Experimental design

The trial was planted in May 1977 at the La Mé Station in the Ivory Coast, in a tertiary sand zone one reclaimed forest land. Each hybrid cross was represented by a total of 24 palms, split into six plots of 4 palms each. Hence, a population is represented by 576 palms (24 crosses × 24 palms). The 4 parental crosses (A × B, A × B′, A′ × B and A × B′) were also included in the trial so as to compare the populations to their crosses. Each parental cross was represented by up to 96 palms split into 24 plots.

Observations and measurements

The individual fresh fruit bunch production (FFB) results from the number of bunches produced (NB) and the mean weight of each bunch (MBW, in kg). Because these traits vary considerably with age,

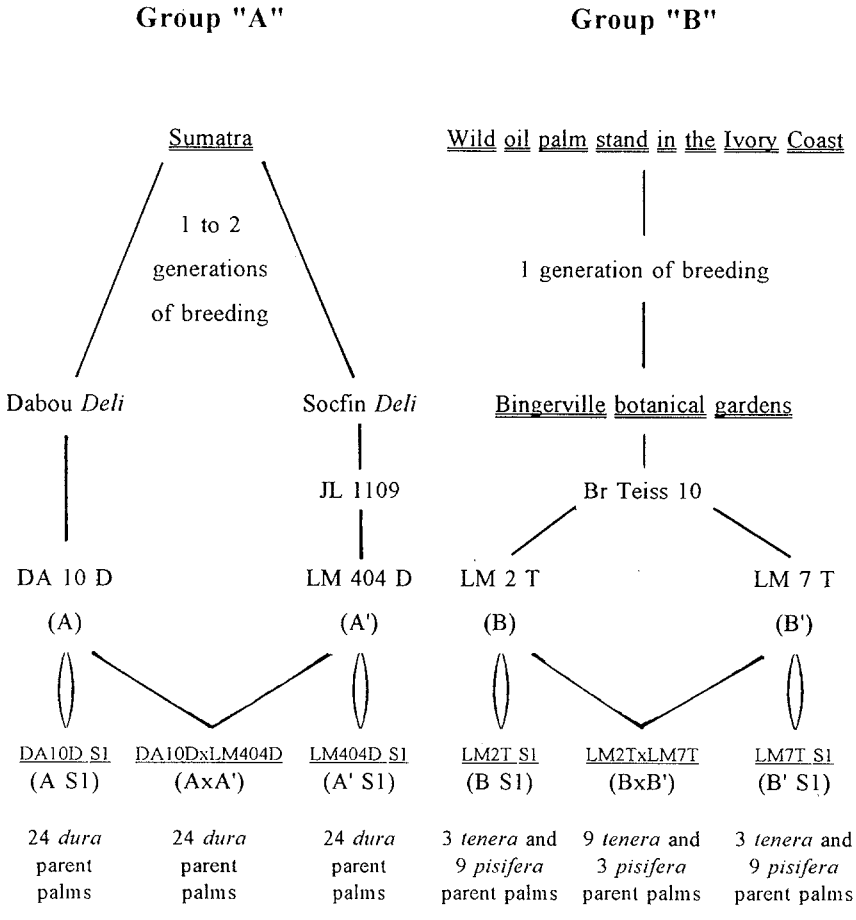

Fig. 1 Pedigree of parent palms and number of individuals per parental family. The original names are *double-underlined*. Family names are *underlined*. Parent names are *not underlined*. The initial palms and their progenies are coded to simplify matters

palms are observed for several years straight from the start of production (year 3) to year 10. Breeders used to consider juvenile and mature stages. In our experiment, the juvenile stage will mean palms 3–6 years old, while the mature stage, those 7–10 years old. In the fields bunches were harvested two or three times a month and the bunch weight recorded. These observations were cumulated at the end of each season, when the annual fresh fruit bunch production and mean bunch weight were computed. The total oil yield (TO, in tons/ha/per year) is calculated from the FFB and the extraction rate (ER, expressed as a percentage of oil weight to bunch weight). The extraction rate was obtained over two analysis campaigns (years 5 and 6) on a sample of bunches from *tenera* palms, the only ones representative of the furture commercial varieties, which are *dura* × *pisifera* ($D \times P$) crosses. It was estimated from three weight ratios: fruit/bunch (FB), mesocarp/fruit (MF) and oil/mesocarp (OM). On average, just over 1 in 3 palms were analysed (37%).

Analysis of variance of the crossing plan

The analysis of variance of the crossing plan is based on the means of full-sib crosses using the following model:

$$G_{abe} = M + F_a + F_b + I_{ab} + R_{abe} \quad (1)$$

G_{abe} = genotypic value of cross e from the P_{ab} population, $e = 1, \ldots 24$;
M = trial mean;
F_a = principal effect of progeny a of *Deli* origin, $a = 0, 1, 2$;
F_b = principal effect of progeny b of *La Mé* origin, $b = 0, 1, 2$;
I_{ab} = deviation of G_{abe} to previous effects;
R_{abe} = within-population cross effect (used as residual in the F-test).

Theory of crossing plan analysis

The statistical approach using the orthogonal contrasts

Using the mothod, we are able to break the total variation between crosses down into as many contrasts as degrees of freedom. The contrasts are defined in the following way:

$$C_k = \sum_{a=0}^{2}\sum_{b=0}^{2}(m_{abk} \cdot \mu_{ab}) \Big/ \sqrt{\sum_{a=0}^{2}\sum_{b=0}^{2} m_{abk}^2}, \text{ where}$$

k varies from 2 to 9. C_1 is the populations mean (i.e. M or $G\ldots$). This convention is adopted for the sake of homogeneity. μ_{ab} is the P_{ab} population mean (i.e. G_{ab}), and m_{abk} is the P_{ab} contribution to C_k. The contributions of each population is given in Table 2. As the contrasts are orthogonal, they can be be tested independently by computing their respective F-ratio: $F_k = 24^* C_k^2 / MS(R_{abe})$, where 24 is the number of crosses in each population. Contrasts are accessible by the experiment. Their genetic significance can be assessed using a genetic model.

Genetic model

The reference population is chosen in such a way that genotypic frequencies are all the same. It may be obtained by a between-group cross between two parental populations in a linkage equilibrium. To do this, a large number of generations of panmixia may be carried out based on A × A′ and B × B′ within-origin crosses. In practice, oil palm breeders are only interested in the third generation (grandchildren), which is the only one tested in the field and corresponds to the move towards varietal creation.

Analysis of individual genotypic value

At a given locus, the alleles of parents A and B can be designed by 1 and 2, and those of A′ and B′ by 3 and 4. There are a maximum of eight different alleles four (four

Table 2 Population contributions (m_{abk}) to the trial mean (C_1) and to contrasts C_2–C_9

Hybrid populations	Trial mean	C_k contrasts							
	C_1	C_2	C_3	C_4	C_5	C_6	C_7	C_8	C_9
P_{00}	1	1	1	1	1	1	1	1	1
P_{01}	1	1	0	0	1	−2	−2	0	−2
P_{02}	1	1	−1	−1	1	1	1	−1	1
P_{10}	1	0	1	0	−2	1	0	−2	−2
P_{11}	1	0	0	0	−2	−2	0	0	4
P_{12}	1	0	−1	0	−2	1	0	2	−2
P_{20}	1	−1	1	−1	1	1	−1	1	1
P_{21}	1	−1	0	0	1	−2	2	0	−2
P_{22}	1	−1	−1	1	1	1	−1	−1	1
Weight coefficients									
$1/\sqrt{\sum_{a=0}^{2}\sum_{b=0}^{2} m_{abk}^2}$	$1/\sqrt{9}$	$1/\sqrt{6}$	$1/\sqrt{6}$	$1/\sqrt{4}$	$1/\sqrt{18}$	$1/\sqrt{18}$	$1/\sqrt{12}$	$1/\sqrt{12}$	$1/\sqrt{28}$

parents × two alleles). Each genotype has one allele from A or A' and one from B or B'. There are 16 elementary genotypes for a single locus and $16^2 = 256$ genotypes for two loci.

Consider two loci, L and M, and let Y_{ijkl} be the value of genotype L_iL_j/M_kM_l. Indices i and k designate the alleles from the *Deli* parents and indices j and l those from the *La Mé* parents. Their values come from the $\{1,2\}$ set if the parents are A or B and from the $\{3,4\}$ set if the parents are A' or B'. According to Kempthorne (1957), Y_{ijkl} can be expressed in the following form:

$$\begin{aligned} Y_{ijkl} &= \mu + \alpha_i + \alpha_j + \alpha_k + \alpha_l \\ &\quad + \beta_{ij} + \beta_{kl} \\ &\quad + (\alpha\alpha)_{ik} + (\alpha\alpha)_{jk} + (\alpha\alpha)_{il} + (\alpha\alpha)_{jl} \\ &\quad + (\alpha\beta)_{ijk} + (\alpha\beta)_{ijl} + (\alpha\beta)_{ikl} + (\alpha\beta)_{jkl} \\ &\quad + (\beta\beta)_{ijkl} \end{aligned} \tag{2}$$

μ = reference population mean; α_i = additive (A) effect of allele L_i; β_{ij} = dominant (D) interaction between homologous alleles; $(\alpha\alpha)_{ik}, (\alpha\alpha)_{il}$ = additive*additive (AA) cis- and trans-interaction between non-homologous alleles; $(\alpha\beta)_{ijk}$ = additive*dominance (AD) interaction between three genes, two of which are homologous; $(\beta\beta)_{ijkl}$ = dominance*dominance (DD) effect between two pairs of homologous alleles.

The effects are independent, each of which (except for the mean) is a random variable whose sum is zero (the summation on a given index produces a nil value, for example $\sum_{i=1}^{4} (\alpha\alpha)_{ik} = 0$).

Analysis of population means

The populations means depend on the allelic and gametic frequencies of the S0 and S1 progenies obtained by crossing and selfing the initial parents:

$$\begin{aligned} \mu_{ab} &= \sum_i \sum_j \sum_k \sum_l p^a_{ik}\, p^b_{jl}\, Y_{ijkl} \\ \mu_{ab} &= \mu + \sum_i p^a_i \alpha_i + \sum_j p^b_j \alpha_j + \sum_k p^a_k \alpha_k + \sum_l p^b_l \alpha_l \\ &\quad + \sum_i \sum_j p^a_i\, p^b_j \beta_{ij} + \sum_k \sum_l p^a_k\, p^b_l \beta_{kl} \\ &\quad + \sum_i \sum_k p^a_{ik} (\alpha\alpha)_{ik} + \sum_k \sum_k p^b_j\, p^a_k (\alpha\alpha)_{jk} \\ &\quad + \sum_i \sum_l p^a_i\, p^b_l (\alpha\alpha)_{il} + + \sum_j \sum_l p^b_{jl} (\alpha\alpha)_{jl} \\ &\quad + \sum_i \sum_j \sum_k p^a_{ik}\, p^b_{ik} (\alpha\beta)_{ijk} + \sum_i \sum_j \sum_l p^b_{jl}\, p^a_i (\alpha\beta)_{ijl} \\ &\quad + \sum_i \sum_j \sum_l p^a_{ik}\, p^b_l (\alpha\beta)_{ikl} + \sum_j \sum_k \sum_l p^b_{jl}\, p^a_k (\alpha\beta)_{jkl} \\ &\quad + \sum_i \sum_j \sum_k \sum_l p^a_{ik}\, p^b_{jl} (\beta\beta)_{ijkl} \end{aligned} \tag{3}$$

p^a_i, p^b_j, p^a_k and p^b_l are allelic frequencies. The index identifies the alleles, while the exponent stands for their origin. Irrespective of the a and b values, the following equalities were checked:

* $p^a_i = p^a_k = p^a$. If a is a selfed progeny, then $p^a = 1/2$, otherwise $p^a = 1/4$;
* $p^b_j = p^b_l = p^b$. If b is a selfed progeny, then $p^b = 1/2$, otherwise $p^b = 1/4$.

p^a_{ik} and p^b_{jl} are gametic frequencies and depend on the recombination rate r and the type of progeny. For the sake of simplicity, r is considered to be 1/2 (no linkage). If so, $p^a_{ik} = 1/2\, p^a$ and $p^b_{jl} = 1/2\, p^b$.

The terms in this model are therefore weighted sums (except for the mean), whose levels depend only on the allelic frequencies. Some factors vary in the same way according to the population because $p^a_i = p^a_k$ and $p^b_j = p^b_l$. So only the exponents, i.e. only the origin of the alleles, need to be considered when examining allelic frequencies. The sums that vary together may be grouped under a common term in the following way:

$$\alpha^A_a = p^a \left[\sum_i \alpha_i + \sum_k \alpha_k \right]$$

$$\alpha^B_b = p^b \left[\sum_j \alpha_j + \sum_l \alpha_l \right]$$

$$\gamma^{AB}_{ab} = p^a p^b \left[\sum_i \sum_j \beta_{ij} + \sum_k \sum_l \beta_{kl} + \sum_i \sum_l (\alpha\alpha)_{il} + \sum_j \sum_k (\alpha\alpha)_{jk} + \frac{1}{4} \sum_i \sum_j \sum_k \sum_l (\beta\beta)_{ijkl} \right]$$

$$\begin{aligned} (\alpha\alpha)^A_a &= \frac{1}{16} \sum_i \sum_k (\alpha\alpha)_{ik} \\ (\alpha\alpha)^B_b &= \frac{1}{16} \sum_j \sum_l (\alpha\alpha)_{jl} \\ (\alpha\beta)^{A/B}_{a/b} &= \frac{1}{16} p^b \left[\sum_i \sum_j \sum_k (\alpha\beta)_{ijk} + \sum_i \sum_k \sum_l (\alpha\beta)_{ikl} \right] \\ (\alpha\beta)^{B/A}_{b/a} &= \frac{1}{16} p^a \left[\sum_i \sum_j \sum_l (\alpha\beta)_{ijl} + \sum_j \sum_k \sum_l (\alpha\beta)_{jkl} \right] \end{aligned} \tag{5}$$

With such conventions,

$$\begin{aligned} \mu_{ab} &= \mu + \alpha^A_a + \alpha^B_b + \gamma^{AB}_{ab} + (\alpha\alpha)^A_a \\ &\quad + (\alpha\alpha)^B_b + (\alpha\beta)^{A/B}_{a/b} + (\alpha\beta)^{B/A}_{b/a} \end{aligned} \tag{6}$$

The exponents still differentiate between origins (*A* for *Deli* and *B* for *La Mé*). Whatever the value of the recombination rate *r*, dominance and AA trans-epistatic effects cannot be estimated independently from the population means. Furthermore, when the recombination rate is 1/2, DD epistatic effects also vary in the same way. Hence, they are grouped in the common γ terms. Only eight independent genetic components can be estimated from population means (Table 3). More details about the calculations are given in the Appendix. The additive, dominance and epistatic effects associated with S0 progenies do not contribute to pupulation means. However, if the loci are not independent ($r < 1/2$), epistatic effects will.

Genetic significance of contrasts

The genetic significance of contrasts is accessed by replacing the μ_{ab} terms with their corresponding combinations of parameters given in Table 3:

$$C_1 = \mu + \frac{1}{3}\left[(\alpha\alpha)_0^A + (\alpha\alpha)_2^A\right] + \frac{1}{3}\left[(\alpha\alpha)_0^B + (\alpha\alpha)_2^B\right]$$

$$C_2 = \alpha_0^A + \frac{1}{2}\left[(\alpha\alpha)_0^A - (\alpha\alpha)_2^A\right] + \frac{1}{3}\left[(\alpha\beta)_{0/0}^{B/A} + (\alpha\beta)_{2/0}^{B/A}\right]$$

$$C_3 = \alpha_0^B + \frac{1}{2}\left[(\alpha\alpha)_0^B - (\alpha\alpha)_2^B\right] + \frac{1}{3}\left[(\alpha\beta)_{0/0}^{A/B} + (\alpha\beta)_{2/0}^{A/B}\right]$$

$$C_4 = \gamma_{00}^{AB} + \frac{1}{2}\left[(\alpha\beta)_{0/0}^{B/A} - (\alpha\beta)_{2/0}^{B/A}\right] + \frac{1}{2}\left[(\alpha\beta)_{0/0}^{A/B} - (\alpha\beta)_{2/0}^{A/B}\right]$$

$$C_5 = \frac{1}{6}\left[(\alpha\alpha)_0^A + (\alpha\alpha)_2^A\right] \qquad (6)$$

$$C_6 = \frac{1}{6}\left[(\alpha\alpha)_0^B + (\alpha\alpha)_2^B\right]$$

$$C_7 = \frac{1}{6}\left[(\alpha\beta)_{0/0}^{B/A} + (\alpha\beta)_{2/0}^{B/A}\right]$$

$$C_8 = \frac{1}{6}\left[(\alpha\beta)_{0/0}^{A/B} + (\alpha\beta)_{2/0}^{A/B}\right]$$

$$C_9 = 0$$

The reference population mean, additive, dominance, AA trans and DD effects only contribute to a single contrast, while AA cis and AD effects contribute up to three contrasts. Contrasts are linear combinations of several genetic effects, if any. Hence C_1, in addition to containing the μ term, also contains AA cis terms. Similarly, C_2 and C_3, corresponding to principal effects, also contain AA *cis* and AD terms. C_4 contains *D*, AA trans, AD and DD terms. C_5–C_8 only contain one effect (AA *cis* for C_5 and C_6, AD for C_7 and C_8). Finally, C_9 does not contain any effect because only two independent loci are considered ($r = 1/2$). If its estimate is significant, then the loci may not segregate independently ($r < 1/2$) or higher-order genetic interactions may be involved. In fact, higher-order genetic effects also affect the other contrasts.

In fact, the relative contribution of genetic effects to contrasts varies according to the recombination rate *r*. The lower the *r* value, the more C_1–C_4 are important. If $r = 0$, then all the effects will be taken into account in C_1–C_4, and the other contrasts will have 0 expectation. The contrasts overestimate the principal effects compared to the interaction effects. On the other hand, the higher the *r* value, the more the estimates of the principal effects will be accurate and the greater the chances of detecting interactions. Robert et al. (1993) showed experimentally and Kervella et al. (1993) by simulation that intermediate to high values of recombination rate (0.196–0.450 in their example) have comparable effects

Table 3 Independent genetic components estimable from the population means, with a recombination rate of 1/2

Population means	Reference population mean	Additive components		Dominance, AA trans and DD epistatic component	AA cis epistatic components		AD epistatic components	
μ_{00}	$= \mu$	$+ \alpha_0^A$	$+ \alpha_0^B$	$+ \gamma_{00}^{AB}$	$+ (\alpha\alpha)_0^A$	$+ (\alpha\alpha)_0^B$	$+ (\alpha\beta)_{0/0}^{B/A}$	$+ (\alpha\beta)_{0/0}^{A/B}$
μ_{01}	$= \mu$	$+ \alpha_0^A$	$+ 0$	$+ 0$	$+ (\alpha\alpha)_0^A$	$+ 0$	$+ 0$	$+ 0$
μ_{02}	$= \mu$	$+ \alpha_0^A$	$+ -\alpha_0^B$	$+ -\gamma_{00}^{AB}$	$+ (\alpha\alpha)_0^A$	$+ (\alpha\alpha)_2^B$	$+ (\alpha\beta)_{2/0}^{B/A}$	$+ -(\alpha\beta)_{0/0}^{A/B}$
μ_{10}	$= \mu$	$+ 0$	$+ \alpha_0^B$	$+ 0$	$+ 0$	$+ (\alpha\alpha)_0^B$	$+ 0$	$+ 0$
μ_{11}	$= \mu$	$+ 0$	$+ 0$	$+ 0$	$+ 0$	$+ 0$	$+ 0$	$+ 0$
μ_{12}	$= \mu$	$+ 0$	$+ -\alpha_0^B$	$+ 0$	$+ 0$	$+ (\alpha\alpha)_2^B$	$+ 0$	$+ 0$
μ_{20}	$= \mu$	$+ -\alpha_0^A$	$+ \alpha_0^B$	$+ -\gamma_{00}^{AB}$	$+ (\alpha\alpha)_2^A$	$+ (\alpha\alpha)_0^B$	$+ -(\alpha\beta)_{0/0}^{B/A}$	$+ (\alpha\beta)_{2/0}^{A/B}$
μ_{21}	$= \mu$	$+ -\alpha_0^A$	$+ 0$	$+ 0$	$+ (\alpha\alpha)_2^A$	$+ 0$	$+ 0$	$+ 0$
μ_{22}	$= \mu$	$+ -\alpha_0^A$	$+ -\alpha_0^B$	$+ \gamma_{00}^{AB}$	$+ (\alpha\alpha)_2^A$	$+ (\alpha\alpha)_2^B$	$+ -(\alpha\beta)_{2/0}^{B/A}$	$+ -(\alpha\beta)_{2/0}^{A/B}$
Number of trms from genetic model (Eq. 2) involved in each component								
16	1	2	2	5	1	1	2	2

on the estimates. Moreover, unless the traits under study are determined by a very small number of linked loci, the average value of r is most likely to be close to 1/2. If so, the P_{11} population mean has an expectancy equal to that of the reference population (Table 3). Furthermore, to estimate the additive effects, AD epistasis has to be nil and AA *cis* contributions to S1 × S1 populations have to be equal (i.e. $(\alpha\alpha)_0^A = (\alpha\alpha)_2^A$ and $(\alpha\alpha)_0^B = (\alpha\alpha)_2^B$) This latter hypothesis is plausible insofar as the parents responsible for these effects share part of their genomes (a common parent for the *La Mé* parents and a common grandparent or great-grandparent for the *Deli* parents). Hence:

$$C_1 = \mu = \mu_{11}$$

$$C_2 = \alpha_0^A \quad \text{and} \quad C_3 = \alpha_0^B$$

$$C_4 = Y_{00}^{AB}$$

$$C_5 = \frac{1}{3}(\alpha\alpha)_0^A \quad \text{and} \quad C_6 = \frac{1}{3}(\alpha\alpha)_0^B$$

Another way of testing the combined effect of overall epistasis and linkage on the mean is to compare the initial crosses ($A \times B$) with the populations derived from them by selfing the parents (AS1 × BS1). In fact, a given cross and the population derived from it contain the same genes, at the same frequencies. They only differ in terms of gametic frequencies. With only two loci, each cross contains four equally frequent genotypes, and the reproduction of the latter results in 16 genotypes, whose frequency depends on the recombination rate.

Experimental results

Comparison of means between the four initial crosses and their S1 × S1 exact reproductions

The LM2T parent (B) transmitted smaller but more numerous bunches than its half-sib LM7T (B′). Its progenies and grand-progenies generally produced more in terms of FFB and oil (Table 4). The DA10D parent (A) transmitted a larger number of bunches than its counterpart LM404D (A′), particularly once adult, but the latter was characterized by heavier, better quality bunches, hence by a better oil yield.

All four S1 × S1 exact reproductions are quite similar to their corresponding parental crosses. The two generations are not different in terms of bunch quality criteria (Table 5). The combined effects of linkage and epistasis on the mean are therefore negligible for these traits, but there is a difference of 1 kg in mean bunch weight that was responsible for a slight reduction (5.5%) in oil yields in the second generation, 3.25 against 3.44 tons/ha per year (Table 6). This result proves the efficiency of the procedure used in the IRHO seed production programme.

When the populations with three grandparents are considered, those obtained from *Deli* S0 plants (i.e. DA10D × LM404D) are intermediate between the corresponding parental crosses, while those obtained from *La Mé* S0 plants (i.e. LM2T × LM7T) produce smaller bunches. This disadvantage is cancelled out when young by a larger number of bunches, but not once an adult. The observed differences may be due to bias linked to

Table 4 Experimental means of the parental crosses for eight traits: number of bunches (NB); mean bunch weight in kg (MBW); fresh fruit bunches in kg/palm per year (FFB); total oil in tons/ha per year (TO); fruit to bunch percentage (FB); mesocarpe to bunch percentage; oil to mesocarpe percentage (OM); and extraction rate (ER)

		LM2T (B)		LM7T (B′)	
		Juvenile	Mature	Juvenile	Mature
DA10D (A)	NB	11.8	9.6	9.6	8.2
	MBW	6.3	12.2	6.6	13.6
	FFB	69.5	109.9	58.6	105.7
	TO	2.01	3.18	1.77	3.20
	FB	60.4		60.7	
	MF	79.0		82.4	
	OM	52.2		52.6	
	ER	25.1		26.2	
LM404D (A′)	NB	11.5	7.5	9.6	6.0
	MBW	6.8	14.9	7.31	16.7
	FFB	71.6	104.8	62.1	95.4
	TO	2.45	3.58	2.06	3.16
	FB	64.0		62.9	
	MF	86.6		86.5	
	OM	53.3		52.7	
	ER	29.7		28.7	

Table 5 Experimental means of the populations for eight traits.[a]

		LM2T SELF		LM2T × LM7T		LM7T SELF	
		Juvenile	Mature	Juvenile	Mature	Juvenile	Mature
	NB	12.5	10.0	12.3	9.01	9.8	8.7
	MBW	6.0	11.5	5.6	11.4	6.3	12.3
	FFB	69.2	113.3	63.3	101.3	56.4	104.5
DA10D SELF	TO	1.94	3.21	1.85	3.02	1.65	3.17
	FB	58.3		60.7		58.0	
	MF	80.5		80.3		82.1	
	OM	51.5		52.0		53.3	
	ER	24.3		25.4		25.4	
	NB	12.0	8.7	11.9	8.3	9.7	7.8
	MBW	6.2	12.6	5.8	11.9	6.4	13.3
	FFB	68.8	107.3	62.3	97.1	56.8	99.6
LM404D × DA10D	TO	2.11	3.35	1.92	3.08	1.81	3.24
	FB	61.7		62.5		61.2	
	MF	82.7		82.1		84.2	
	OM	52.0		51.9		53.4	
	ER	26.6		26.7		27.5	
	NB	11.2	7.8	11.7	7.4	9.2	6.8
	MBW	6.3	13.5	5.9	12.8	6.7	14.3
	FFB	64.2	102.3	61.7	92.5	55.7	93.1
LM404D SELF	TO	2.11	3.42	2.07	3.18	1.86	3.20
	FB	64.6		65.5		62.4	
	MF	85.8		85.0		86.9	
	OM	51.4		52.0		53.2	
	ER	28.5		29.0		28.9	

[a] See Table 4 for explanation of abbreviations

Table 6 Comparison between the initial crosses and their exact reproductions for eight traits.[a] Only the mature stage is considered

	Traits							
	NB	MBW	FFB	FB	MF	OM	ER	TO
Mean of initial crosses[b]	7.8	13.9	104.1	61.9	83.6	52.7	27.4	3.44
Mean of exact reproductions[c]	8.3	12.9	103.3	60.8	83.8	52.3	26.8	3.25
F-test[d]	NS	**	NS	NS	NS	NS	NS	*

[a] See Table 4 for explanation of abbreviations
[b] The initial crosses are A × B, A′ × B, A × B′ and A′ × B′
[c] Their exact reproductions: A S1 × B S1, A′ S1 × B S1, A S1 × B′ S1 and A′ S1 × B S1
[d] The F-test is based on cross means.
*** $P \leq 0.001$; ** $0.001 < P \leq 0.01$; * $0.01 < P \leq 0.05$; NS, $P > 0.05$

the choice of parents, since they were chosen on their intrinsic merits for highly heritable traits (MF, OM, NB).

Although the choice is made in the same way for all three *Deli* progenies, it can be different for *La Mé* progenies, because of the fruit shell thickness. In effect, contrary to *dura* and *tenera* genotypes, *pisifera* genotypes cannot be chosen according to their intrinsic merits since they are sterile. The relative excess of *tenera* in the *La Mé* S0 progeny suggests that the selection pressure applied to this progeny was greater than the one applied on the selves. In practice, this selection was largely ineffective for yield, since the *pisifera* have the same combining ability as their *tenera* full-sibs (test not shown).

Analysis of variance and of contrasts

Depending on the traits considered, the ancestry of the parents accounts for 11% (TO) to 56% (MF) of the variation between hybrid crosses (Table 7). That is to

Table 7 Analysis of sources of variation between hybrid crosses of all populations and for eight traits[a]. Only the mature stage is considered

Source of variation		*df*[c]	Mean squares for several traits[d]							
			NB	MBW	FFB	FB	MF	OM	ER	TO
Model		8	22.2 ***	21.8 ***	1073.5 ***	150.1 ***	129.5 ***	13.8 **	68.8 ***	0.356 **
$F_a(C_2+C_5)$		2	66.6 ***	57.2 ***	1955.0 ***	477.8 ***	445.6 ***	0.8 NS	257.9 ***	0.337 NS
	C_2	1	133.2 ***	114.3 ***	3908.9 ***	953.8 ***	883.4 ***	0.0 NS	515.7 ***	0.653 ***
	C_5	1	0.0 NS	0.0 NS	1.0 NS	1.8 NS	7.8 NS	1.6 NS	0.0 NS	0.022 NS
$F_b(C_3+C_6)$		2	20.9 ***	28.3 ***	2308.1 ***	101.1 ***	71.3 ***	53.0 ***	12.4 *	0.964 ***
	C_3	1	41.8 ***	21.3 ***	2676.9 ***	37.0 NS	70.0 ***	95.2 ***	23.7 ***	0.569 **
	C_6	1	0.0 NS	35.4 ***	1939.4 ***	165.0 ***	72.6 ***	10.8 *	1.2 NS	1.359 ***
$I_{ab}(C_4+C_7+C_8+C_9)$		4	0.6 NS	0.8 NS	15.1 NS	10.5 NS	0.6 NS	0.7 NS	2.5 NS	0.062 NS
R_{abc}(residual)		206	1.1	1.7	62.3	11.1	3.96	2.1	3.4	0.114
R^{2e}			0.45	0.33	0.40	0.35	0.56	0.20	0.44	0.11

[a] See Table 4 for explanation of abbreviations

[b] F_a, F_b, I_{ab}, and R_{abe} are effects defined in Eq. 1. Means squares of the contrasts that are related to each of these effects are expressed directly below, except for I_{ab} (see their definitions in Eq. 7)

[c] The degree of freedom of the residual is not equal to 9*23 as expected because 1 cross was missing. Only 215 crosses were available instead of 216 (24*9)

[d] Significance levels of *F*-tests:*** ($P \leqslant 0.001$), **($0.001 < P \leqslant 0.01$), *($0.01 < P \leqslant 0.05$) and NS ($P > 0.5$)

[e] R^2 is the determination coefficient of the model of the analysis of variance based on the cross means (G_{abe}):

$$R^2 = \sum_{a=0}^{2}\sum_{b=0}^{2}(G_{ab}-G_{...}) \Big/ \sum_{a=0}^{2}\sum_{b=0}^{2}(G_{abe}-G_{...})^2$$

say, a great part of the variation is due to differences between parents within S0 or S1 progenies, in particular for the oil yield (TO), which is the most complex and the least heritable trait but the one that interests most breeders. This is not surprising because oil palm is a allogamous species. Therefore, grandparents, are very likely to be heterozygous at most of the loci involved in the observed variation.

The variation arises from within *Deli* progenies as well as from those of *La Mé* origin, except for oil to mesocarp, ratio, for which the *Deli* do not differ and for oil production, for which they differ only at the $P = 0.053$ level. With regard to the other traits, *Deli* progenies differ even more than their *La Mé* counterparts, except for FFB. The F_a effect mean squares is twice (MBW) to 20 times (ER) as great as those of the F_b effect.

On the other hand, the interaction between families (I_{ab}) is not significant for any trait. Neither are individual contrasts that are associated with it (C_4, C_7, C_8 and C_9– data not shown). So, dominance and epistatic componets, apart from AA cis, probably do not contribute to population means. As a result, the remaining contrasts, which are associated with F_a and F_b principal effects, are very likely to be good estimators of both additive and AA cis components. These contrasts reveal a real asymmetry between origins. Principal or additive effects are more significant in the *Deli* (C_2) than in the *La Mé* (C_3) origin, even for oil yield. C_2 is not significant only for the oil to mesocarp ratio already mentioned above. As for AA cis epistatic effects, they are substantial in the *La Mé* (C_6) origin, but are never significant ($P > 0.05$) in the *Deli* (C_5) origin.

The low value of hybrid crosses from (LM2T × LM7T) progeny compared to those of the corresponding selfed progenies is an AA cis epistatic effect within the *La Mé* origin. This effect is at least as great as the *La Mé* additive effect for five out of eight traits studied (MBW, FFB, FB, MB and TO). Among them, FFB, FB and TO are known as the least heritable by many oil palm breeders, (Soh and Tan 1981).

Discussion

Representativeness of the material

The grandparents in our experiment are very widely used both for seed production and further breeding work, hence the majority of the varieties currently distributed by the stations associated with CIRAD-CP (ex-IRHO) have LM2T as a parent (from 25% to 50% of their genomes). The parents are typical of their origins: LM2T and LM7T very likely come from an open-pollinated wild palm, high heterozygosity and low divergence are therefore to be expected. On the other hand, the *Deli* populations can be traced back to four seeds, probably from the same mother palm, imported into Indonesia at the start of the 19th century (Hartley 1988). LM404D and DA10D, obtained from two of these populations, are separated from their shared African ancestor by several generations of independent mass selection, three for DA10D and five for LM404D (Fig. 1). Greater homozygosity and greater genetic divergence are therefore to be expected.

Relations between genetic effects and the origin and structure of the material

Although the poor performance of the (LM2T × LM7T) family may be partly due to sampling bias, its low genetic value, probably resulting from poor complementation between the two parent palms, is primarily to blame. This lack of complementation can be

interpreted as a AA cis epistatic effect. In fact, the existence of additive*additive epistatis seems to be common in oil palm breeding, and may be expressed favourably or unfavourably. Cases of favourable recombination have been very often observed between *Deli* parents (Cochard et al. 1993), very likely since they were never crossed at random but crossed on the basis of their complementary characteristics. Hence, DA10D was selected for its prolificacy and LM404D, for its high bunch quality. This complementation phenomenon would be favoured by the fixation beforehand of the genes responsible for the observed traits and explains why it is less often seen in recombinations between highly heterozygous structures of African origins such as *La Mé*, which has been subjected to less selection pressure.

Implications for oil palm breeding

When breeding a perennial species such as oil palm, a breeding cycle is followed by a single generation of intercrossing with a low number of individuals. Assortative mating between complementary parents is often favoured, so as to more rapidly increase the frequency of favourable alleles in the next cycle of progeny tests. As a result, linkage disequilibrium and epistasis probably play a predominant role. In particular, our experimental results showed that epistasis may be more important than dominance. Yet it should be kept in mind that dominance is very likely underestimated because gene effects are defined according to two distinct parent populations instead of an unique reference population (Stuber and Cockerham 1966). Even then epistasis is sometimes more important than additive effects. In oil palm, some epistatic effects have been widely used in breeding and varietal creation. This involves using information already available about the parents, such as geographical origin, inbreeding, relatedness or complementary features.

In fact, trials like the one studied in this article are difficult to carry out when the number of parents increases. Yet, the AA cis component of epistasis may be easily estimated in regular crossing plans: one has only to compare families such as A self, A′ self and A × A′ cross using a common tester population. Such families are usually carried out in oil palm breeding programmes. A × A′ type families are the basic material of the next cycle of the breeding scheme, while A and A′ selfed families are used to produce commercial seeds and sometimes to further progeny tests.

In practice, when progeny-testing palms from A × A′ type families, two supplementary check crosses from the original parents or their respective selfed progenies should be included in the trial. Even if parents are more or less inbred, their hybrid crosses to be progeny tested are not. So genetic values may be estimated with the same accuracy, contrary to the *A*, *B* or *C* methods of Mather, which imply completely inbred lines, F_1 and F_2 or backcross generations (Mather 1949). The O-test on the F_2 (Opsahl 1956), also called the triple testcross by Jinks (1983) or T-test on backcrosses are more accurate because all generations are collateral (Van der Veen 1959). O and T contrasts are quite similar to our C_5 and C_6 although they have not the same genetic significance since they are not the same linear combinations of genetic effects (Hayman and Mather 1955). Yet, the O contrast detects AA epistasis, but it does not attribute the effect to one group or the other. Lastly, our C_5 and C_6 contrasts correspond most to Bauman's (1959) or Mather and Jinks' schemes (1971).

Conclusion

Epistasis effects may contribute substantially to population means if the material tested is highly heterozygous, the genetic base is narrow (selected material or few individuals used) or there is linkage disequilibrium (due to further selection and insufficient intercrossing generations). All these occur in oil palm and prompted the setting up of the experiment described in this article.

The design studied is suitable for analysing the different components of digenic epistasis. Within a reciprocal recurrent selection scheme, it can be used to determine the components that can be attributed to each of the two heterotic groups. It involves collaterals whose values are known with equal accuracy (single experimental protocol). Parents may have any degree of inbreeding. The design makes it possible to isolate the AA component, which can be used for varietal creation, from the other components that cannot.

Appendix

Expression of genetic components involved in population means, with a recombination rate of 1/2

–The additive component associated with each origin has three levels. For example, in the *Deli* origin, the levels are α_0^A, α_2^A, and α_1^A, corresponding to A self and A′ self and A × A′ cross families, respectively:

$$\alpha_0^A = \sum_{i=1}^{2} p_i^0 \alpha_i + \sum_{k=1}^{2} p_k^0 \alpha_k = \frac{1}{2}\left[\sum_{i=1}^{2} \alpha_i + \sum_{k=1}^{2} \alpha_k\right]$$

$$\alpha_2^A = \sum_{i=3}^{4} p_i^2 \alpha_i + \sum_{k=3}^{4} p_k^2 \alpha_k = \frac{1}{2}\left[\sum_{i=3}^{4} \alpha_i + \sum_{k=3}^{4} \alpha_k\right] = -\alpha_0^A$$

$$\alpha_1^A = \sum_{i=1}^{4} p_i^1 \alpha_i + \sum_{k=1}^{4} p_k^1 \alpha_k = \frac{1}{4}\left[\sum_{i=1}^{4} \alpha_i + \sum_{k=1}^{4} \alpha_k\right] = 0$$

The additive component associated with the *La Mé* origin can be estimated in the same way (not shown).

–The common component to dominance, AA trans and DD epistasis takes values as follows:

$$\gamma_{00}^{AB} = \frac{1}{4}\left[\sum_{i=1}^{2}\sum_{j=1}^{2} \beta_{ij} + \sum_{k=1}^{2}\sum_{l=1}^{2} \beta_{kl} + \sum_{i=1}^{2}\sum_{l=1}^{2} (\alpha\alpha)_{il} + \sum_{j=1}^{2}\sum_{k=1}^{2} (\alpha\alpha)_{jk} + \frac{1}{4}\sum_{i=1}^{2}\sum_{j=1}^{2}\sum_{k=1}^{2}\sum_{l=1}^{2} (\beta\beta)_{ijkl}\right] = \gamma_{22}^{AB}$$

$$\gamma_{02}^{AB}=\frac{1}{4}\left[\sum_{i=1}^{2}\sum_{j=3}^{4}\beta_{ij}+\sum_{k=1}^{2}\sum_{l=3}^{4}\beta_{kl}+\sum_{i=1}^{2}\sum_{l=3}^{4}(\alpha\alpha)_{il}+\sum_{j=3}^{4}\sum_{k=1}^{2}(\alpha\alpha)_{jk}\right.$$
$$\left.+\frac{1}{4}\sum_{i=1}^{2}\sum_{j=3}^{4}\sum_{k=1}^{2}\sum_{l=3}^{4}(\beta\beta)_{ijkl}\right]=\gamma_{20}^{AB}=-\gamma_{00}^{AB}$$

$$\gamma_{01}^{AB}=\frac{1}{8}\left[\sum_{i=1}^{2}\sum_{j=1}^{4}\beta_{ij}+\sum_{k=1}^{2}\sum_{l=1}^{4}\beta_{kl}+\sum_{i=1}^{2}\sum_{l=1}^{4}(\alpha\alpha)_{il}+\sum_{j=1}^{4}\sum_{k=1}^{2}(\alpha\alpha)_{jk}\right.$$
$$\left.+\frac{1}{4}\sum_{i=1}^{2}\sum_{j=1}^{4}\sum_{k=1}^{2}\sum_{l=1}^{4}(\beta\beta)_{ijkl}\right]=\gamma_{21}^{AB}=\gamma_{10}^{AB}=\gamma_{12}^{AB}=\gamma_{11}^{AB}=0$$

–The three levels AA cis epistasis, associated with the *Deli* origin, are expressed as follows:

$$(\alpha\alpha)_0^A=\sum_{i=1}^{2}\sum_{k=1}^{2}p_{ik}^0(\alpha\alpha)_{ik}=\frac{1}{4}\left[\sum_{i=1}^{2}\sum_{k=1}^{2}(\alpha\alpha)_{ik}\right]$$

$$(\alpha\alpha)_2^A=\sum_{i=3}^{4}\sum_{k=3}^{4}p_{ik}^2(\alpha\alpha)_{ik}=\frac{1}{4}\left[\sum_{i=3}^{4}\sum_{k=3}^{4}(\alpha\alpha)_{ik}\right]$$

$$(\alpha\alpha)_1^A=\sum_{i=1}^{4}\sum_{k=1}^{4}p_{ik}^1(\alpha\alpha)_{ik}=\frac{1}{16}\left[\sum_{i=1}^{4}\sum_{k=1}^{4}(\alpha\alpha)_{ik}\right]=0$$

The same can be done for the *La Mé* origin (not shown).

–The levels of AD component depend on both the parents and the testers. For example, the levels related to *Deli* familes, tested on *B* self family can be expressed as:

$$(\alpha\beta)_{0/0}^{A/B}=p^0\left[\sum_{i=1}^{2}\sum_{j=1}^{2}\sum_{k=1}^{2}p_{ik}^0(\alpha\beta)_{ijk}+\sum_{i=1}^{2}\sum_{k=1}^{2}\sum_{l=1}^{2}p_{ik}^0(\alpha\beta)_{ikl}\right]$$
$$=\frac{1}{8}\left[\sum_{i=1}^{2}\sum_{j=1}^{2}\sum_{k=1}^{2}(\alpha\beta)_{ijk}+\sum_{i=1}^{2}\sum_{k=1}^{2}\sum_{l=1}^{2}(\alpha\beta)_{ikl}\right]$$

$$(\alpha\beta)_{2/0}^{A/B}=p^0\left[\sum_{i=3}^{4}\sum_{j=1}^{2}\sum_{k=3}^{4}p_{ik}^2(\alpha\beta)_{ijk}+\sum_{i=3}^{4}\sum_{k=3}^{4}\sum_{l=1}^{2}p_{ik}^2(\alpha\beta)_{ikl}\right]$$
$$=\frac{1}{8}\left[\sum_{i=3}^{4}\sum_{j=1}^{2}\sum_{k=3}^{4}(\alpha\beta)_{ijk}+\sum_{i=3}^{4}\sum_{k=3}^{4}\sum_{l=1}^{2}(\alpha\beta)_{ikl}\right]$$

$$(\alpha\beta)_{1/0}^{A/B}=p^0\left[\sum_{i=1}^{4}\sum_{j=1}^{2}\sum_{k=1}^{4}p_{ik}^1(\alpha\beta)_{ijk}+\sum_{i=1}^{4}\sum_{k=1}^{2}\sum_{l=1}^{2}p_{ik}^1(\alpha\beta)_{ikl}\right]=0$$

If the testers are from B′ self, the corresponding values are the inverse of those associated with B self. If the testers are from the B × B′ cross, the values become nil. Therefore, there are five distinct levels in AD components. The same can be done for $(\alpha\beta)^{B/A}$ terms (not shown).

References

Bauman LF (1959) Evidence of non-allelic gene interaction in determining yield, ear height and kernel row number in corn. Agron J 51:531–534

Beirnaert A, Vanderweyen R (19410 Contribution àl'étude génétique et biométrique des variétés d'*Elaeis guineensis* Jacquin. Publications de l'I.N.E.A.C., Séries Scientifiques no. 27

Cochard B, Noiret JM, Baudouin L, Flori, A, Amblard Ph (1993) Second-cycle reciprocal recurrent selection in oil palm (*Elaeis guineensis* Jacq.): results of *Deli* × *La Mé* hybrids tests. Oléagineux 48:441–451

Gascon JP, De Berchoux C (1963) Quelques relations entre les *dura* et *tenera* d'une même descendance et leur application à l'amélioration des semences, Caractéristiques qualitatives des régimes d'*Elaeis guineensis* Jacq. Oléagineux 18:411–415

Hartley CWS (1988) The oil palm (*Elacis guineenis* Jacq.). 3rd edn. Longman scientific and technical, London

Hayman BI, Mather K (1955) The description of genetic interaction in continuous variation. Biometrics 11:69–82

Jacquemard JC, Meunier J, Bonnot F (1981) Etude génétique de la reproduction d'un croisement chez le plamier à huile (*Elaeis guineensis* Jacq.). Oléagineux 36:343–350

Jinks JL (1983) Biometrical genetics of heterosis. In: Frankel H (eds) Heterosis-Reapraisal of theory and practice. Springer, Berlin Heidelberg New York, pp 1–46

Kempthorne O (1957) An introduction to genetic statistics. John Wiley & Sons, New York; Chapman & Hall, London

Kervella J, Robert N, Fouilloux G (1993) Influence de la recombinaison sur la variabilité génétique. II. Etude par simulations. Agronomie 13:371–379

Mather K, Jinks JL (1971) Biometrical genetics. The study of continuous variation, 2nd edn. Chapman and Hall, London

Meunier J, Gascon JP (1972) Le schéma général d'amélioration du palmier à huile à l'IRHO. Oléagineux 27:1–12

Opsahl B (1956) The discrimination of interaction and linkage in continuous variation. Biometrics. 10:415–432

Robert N, Kervella J, Fouilloux G (1993) Influence de la recombinaison sur la variabilité génétique. I. Etude expérimentale. Agronomie 13:275–281

Soh AC, Tan ST (1981) Estimation of genetic variance, heritability and combining ability in oil palm breeding. In: Yap TC, Graham KM, Sukaimi J (eds) Crop improvement research. Proc 4th Int SABRAO Cong. The society for the Advancement of Breeding Researches in Asia and Oceania, Kuala Lumpur, pp 379–388

Stuber CW, Cockerham CC (1966) Gene effects and variances in hybrid populations. Genetics 54:1279–1286

Van der Veen JN (1959) Test on non-allelic interaction and linkage for quantitative characters in generations derived from two diploid pure lines. Genetics 30:201–232