

# The use of molecular markers to investigate the genetic structure of an oil palm breeding programme

S. MAYES†, P. L. JACK\*‡ & R. H. V. CORLEY§

†Department of Genetics, University of Cambridge, The Downing Site, Cambridge, CB3 3EF, U.K.,

‡Monsanto Cambridge, Maris Lane, Trumpington, Cambridge, CB2 2LQ, U.K. and §Highlands, New Road, Great Barford, Bedford, MK44 3LQ, U.K.

RFLP markers (40 probes covering 60% of the oil palm genome) have been used to assess genetic diversity within 54 palms of a specific oil palm breeding programme. A further 10 palms encompassing a broader range of origins were also included to provide a wider framework for comparative analysis. These palms represent a majority of the parents in a crossing programme which aims to combine the best features of African tenera germplasm with South-east Asian Deli dura material. Progeny from these crosses are planted in up to six sites on three continents, where the palms are being assessed in detail for yield components. A total of 157 RFLP bands were scored and the data analysed by calculating genetic distances according to Nei & Li, and by correspondence and cluster analysis. The relationships that emerge through this molecular analysis correspond well with known pedigree and provenance. For example, south-east Asian Deli dura material forms a cluster clearly distinct from AVROS pisifera gene pools, and a selection of African breeding material forms a further and broader grouping. The potential ability of markers to assist in oil palm breeding through examination of the genetic structure of crossing programmes is discussed, as is their value in parental selection to maintain residual diversity within specific breeding pools and use in evaluating the fidelity of breeding lineages.

**Keywords:** *Elaeis guineensis*, genetic diversity, oil palm, RFLP.

## Introduction

Oil palm (*Elaeis guineensis*) now accounts for approximately 25% of world production of vegetable oil, second only to soybean at 29%. In terms of yield of oil per hectare, the oil palm is by far the most productive oil crop, with the best palms in Malaysia regularly producing six tonnes of oil/hectare/year, exceeding by a factor of at least four that from temperate crops such as soybean and oil seed rape.

Approximately five generations of selection and breeding have been completed in oil palm since the establishment of the first plantations in the 1920s and 1930s (Hartley, 1988). The rate is primarily limited by the long generation and selection cycles, typically 12 years. Despite this, oil production has increased fourfold in Malaysia in the last 50 years, half of this being attributable to genetic improvement of material (Corley & Lee, 1992). However, a recent study has suggested that selection for this improvement has

reduced the coefficient of variation for fruit bunch yield by nearly half (Lee *et al.*, 1990). In recent years, concern has grown that intensive breeding from a narrow genetic base, particularly in the south-east Asian industry, is leading to limited variation within elite material, potentially reducing the rate of future breeding progress (Rosenquist, 1986). An objective measure of genetic variation would allow the selective use of wild and semiwild genotypes and material derived from other commercial sources to widen the genetic base in a precise manner. Until recently, the only information linking different breeding populations was pedigree, inadequate in many cases and sometimes incorrect.

In recent years the use of molecular markers to assess diversity within germplasm collections using isozyme (Ghesquiere, 1984; Baudouin, 1992) and RAPD (Shah *et al.*, 1994) systems has been addressed. Previously we have reported the use of RFLP markers to distinguish genotypes, both inter- and intraspecifically (Jack *et al.*, 1995; Mayes *et al.*, 1996) and have reported the development of a RFLP linkage map for oil palm (Mayes *et al.*, 1997). We describe here the use of 40 probes from

\*Correspondence. E-mail: peter.jack@monsanto.com

this map, covering 60% of the oil palm genome, to assess RFLP variability in a representative sample of parents derived from one major breeding programme. This programme (Combined Breeding Programme 2 — CBP2) aims to combined the best characters from African and south-east Asian material.

## Materials and methods

### Genetic material

Combined Breeding Programme 2 was planted between 1987 and 1989 and consists of crosses between elite material from breeding programmes at Dami, Papua New Guinea (NBPOL) and at Binga, Democratic Republic of Congo (DRC). Crosses were planted in up to six locations, primarily: Binga and Yaligimba (DRC), Dami (Papua New Guinea), Bah Lias (Indonesia), RISPA (Indonesia), Univanich (Thailand) and Unipalma (Colombia). In addition to the 54 samples from CBP2, 10 other palm samples were included in the analysis as follows: six AVROS (Algemene Vereniging van Rubber-planters ten Oostkust van Sumatra) samples, consisting of five samples from the self-pollination of palm SP540T and the palm AP1540.39. In addition, palm 312/3 (the parent self-pollinated to give progeny Binga 143) and also three accessions from a 1984 collection of Tanzanian wild palms (Rajanaidu, 1987), which were included to act as outliers and to help establish their relationship to other African palms. Further information concerning provenance is given in Table 1. Leaf samples (frond 17) were air-freighted from plantations and stored at  $-70^{\circ}\text{C}$  before DNA extraction.

### RFLP techniques

DNA extraction, Southern blotting, probe preparation and autoradiography were all as previously described (Jack *et al.*, 1995). DNA from the complete set of palm genotypes was digested with one of four endonucleases (*EcoRI*, *EcoRV*, *HindIII* and *SstI*). Forty probes were chosen on the basis of their mapped position (Mayes *et al.*, 1997), to give an even sampling of the oil palm genome. From the mapping data, the probes chosen are likely to give a minimum coverage of 60% of the oil palm genome.

### Data analysis

This was carried out using procedures which have now become standard for phenetic analysis of RFLP and other DNA marker data (Virk *et al.*, 1995; Xiao *et al.*, 1996). For each probe/enzyme combination, autoradiographs were scored for the presence or absence of bands

**Table 1** Provenance of oil palm material analysed by RFLP

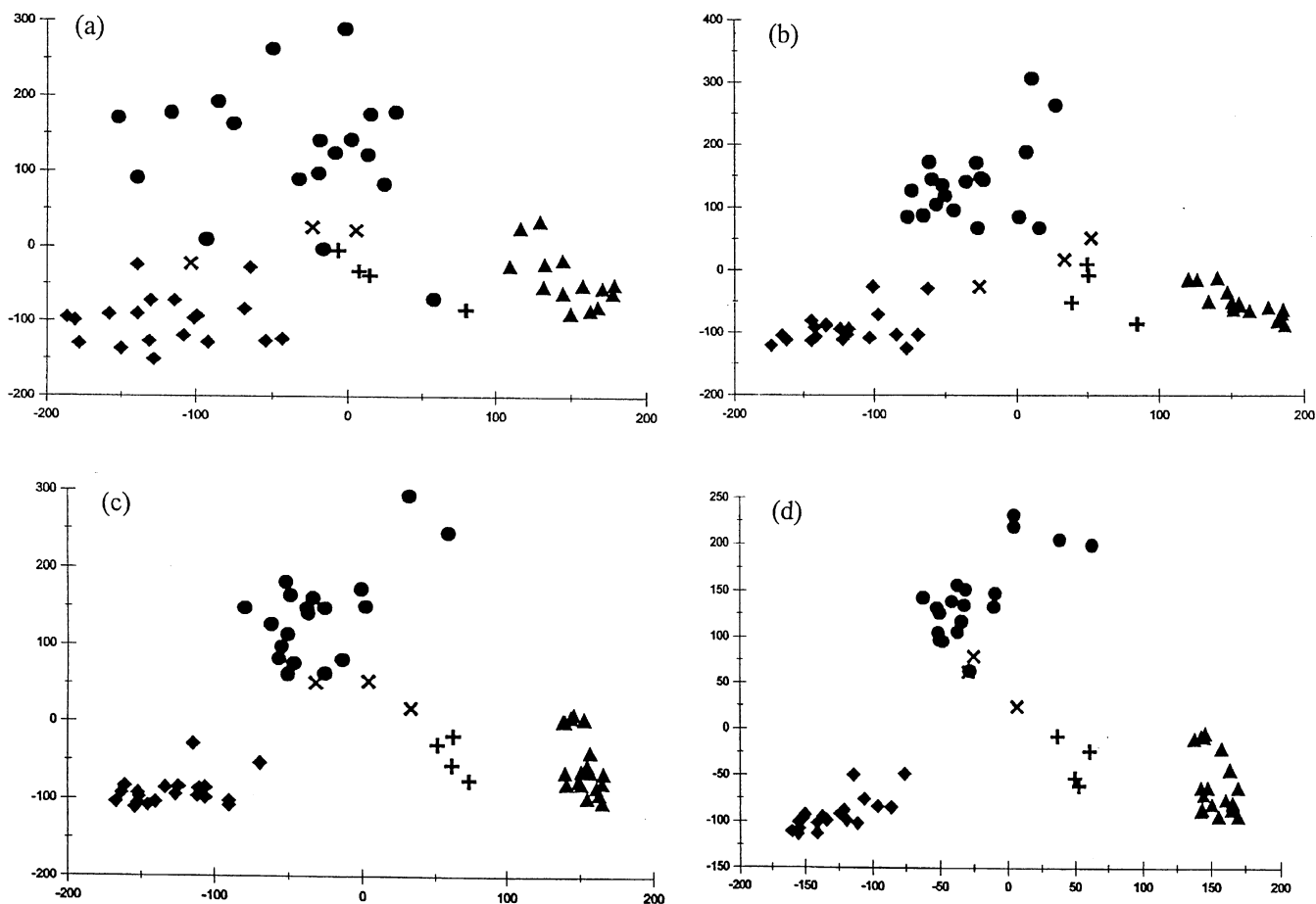
Progeny descriptor	Origin or population type	Number of palms examined	Symbol in Fig. 1
Binga 312/3	DRC	1	●
Binga 30	DRC × Nigeria	2	●
Binga 142	DRC	5	●
Binga 143	DRC	2	●
Binga 271	75% DRC/25% Deli	3	●
Binga 414	75% DRC/25% Deli	6	●
Dm 703	Dami dura	5	◆
Dm 711	Dami dura	6	◆
Dm 712	Dami dura	2	◆
Dm 714	Dami × Ulu Remis dura	5	◆
Dm 715	Dami × Ulu Remis dura	1	◆
Dm 742	AVROS pisifera	8	▲
Dm 568	AVROS pisifera	5	▲
SP540 F <sub>2</sub>	AVROS parental	5	▲
AP1540	AVROS parental	1	▲
Dm 603	(Deli × Serdang p.) × AVROS	3	+
Dm 735	Deli dura × AVROS	1	+
Tanzanian	Tanzanian Collection	3	x

observed across the set of 64 palms. From 40 probes, a total of 157 bands were scored. This binary matrix was analysed by calculating genetic distances (Nei & Li, 1979), correspondence analysis and cluster analysis (Nearest-Neighbours method) using the Multivariate Statistical Package (mvsp v. 2.1, Kovach, 1993).

## Results

In order to identify how many probes might be needed to gain a robust measure of relationships, data were analysed in turn from 10, 20, 30 or 40 probe hybridizations using correspondence analysis (Fig. 1a–d). With 10 probes, three main groups of material begin to separate. With 20 probes, this clustering of main groups becomes more pronounced and it was possible to identify each palm uniquely despite considerable similarities in the pedigree of some of the progenies. With data from 40 probes, the relative positions of material are very similar to the results from 30 probes, suggesting that the distribution has stabilized and hence reflects a true picture of the genetic relationships. The analysis was repeated a number of times taking different subsets of the data and similar results were obtained, suggesting that the data set is not biased by a particular subset of probes.

Figure 1(d) (40 probes) identifies separation of material into groups consistent with their origins. For example, the AVROS pisifera population is derived



**Fig. 1** Correspondence analysis for the 64 oil palms studied, using data obtained with: (a) 10 (b) 20 (c) 30 and (d) 40 RFLP probes. The percentage variation accounted for in axis 1 (horizontal axis) and axis 2 (vertical axis) is: (a) 21% and 14% (b) 23% and 12% (c) 21% and 10% and (d) 20% and 11%. Different palm progenies/origins are identified by different symbols as described in Table 1.

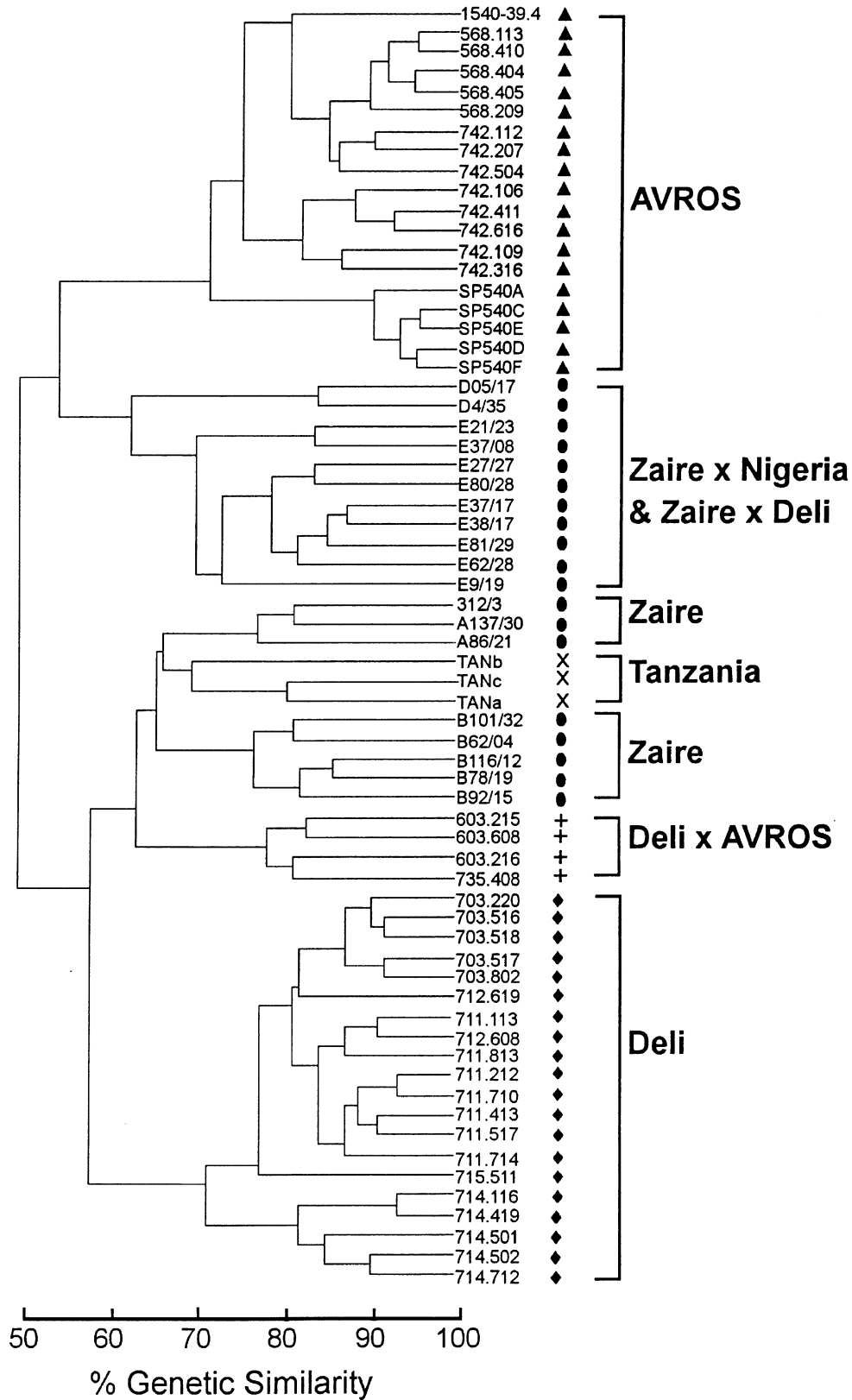
from two main sources of genes, namely palms SP540T and Pol.3184P and all 19 AVROS-related palms cluster at the bottom right of Fig. 1(d) (▲). In contrast, the Dami and Dami × Ulu *Remis dura* populations, originally derived from four Deli *dura* palms planted in the Bogor Botanical Garden in 1848, cluster at the bottom left (◆) of Fig. 1(d). The four *dura* × *pisifera* hybrids are at an intermediate position between these two clusters (+). The 19 palms with a DRC origin form a further distinct group (middle/top of Fig. 1d, ●).

The dataset was also examined by cluster analysis as shown in Fig. 2. Four major clusters can be identified as follows. The top cluster consists solely, and includes all, pure AVROS palms studied. In contrast the bottom cluster consists only, and includes all, pure Deli types. The two rather less discrete clusters in the centre of the figure consist either of DRC origins crossed with Nigerian/Deli material, or of pure DRC or pure Tanzanian or Deli × AVROS crosses. The AVROS

material shows the least intragroup variability, followed by the Deli types. Analysis using cumulative data to examine how the cluster analysis evolves as more information is added, suggests that the progeny groups and the AVROS and Deli populations form rapidly, but the exact relationship between the African groups and the AVROS and Deli groups remains relatively fluid (data not shown). This may in part be explained by the much wider genetic base present in the DRC material, compared to the AVROS and the Deli material. These observations are very much in line with those obtained in the correspondence analysis described above.

## Discussion

Oil palm has one well-characterized Mendelian gene of overwhelming agronomic importance, that of shell-thickness (Hartley, 1988). Oil yield of the heterozygous *tenera* form (*sh*<sup>+</sup>/*sh*<sup>-</sup>) is at least 30% higher (Corley &



**Fig. 2** Cluster analysis for the entire data set (64 palms and 40 RFLP probes). Symbols are as in Table 1. Individual palm genotypes are given as well as a more generic descriptor of origins.

Lee, 1992) than the thick-shelled homozygous dura form ( $sh+/sh+$ ), with the shell-less homozygous pisifera form ( $sh-/sh-$ ) often producing no yield because of female sterility. Almost all commercial material planted today is of the tenera, thin-shelled, phenotype. This necessitates the development of separate pools of pisifera and dura parents to produce hybrid tenera planting material. The Deli dura material, which is predominant in south-east Asia, is derived from four palms planted in the Botanical garden at Bogor in Indonesia in 1848. When, in the 1950s and 1960s, it became clear that tenera fruit had major yield advantages, the industries in Malaysia and Indonesia began major expansion programmes based upon tenera types. To produce this material, pisifera pollen parents had to be developed and again these were often derived from extremely narrow genetic bases. In the case of the AVROS population, most of the genes currently in use are derived from only two palms (Pol3184P and SP540T). The descendants of these palms dominate the pisifera gene pools in south-east Asia. The concern that the variation remaining in these pools will limit future breeding progress is wide-spread in south-east Asia and numerous programmes have begun to introduce material from wider sources. There are a number of ways in which this might be achieved. For example, different breeding programmes may apply different selection criteria. Thus elite material from different plantations, but the same founding population, may allow some widening of the genetic base without loss of breeding progress. An example of this might be seen in Fig. 2 where Deli germplasm splits into two subgroups. Although all of this material ultimately derives from the same origin, it has undergone improvement and selection in different locations (Rosenquist, 1986). Thus one subgroup (DM 703, DM 711, DM 712) contains material predominantly derived from the 'Serdang Avenue' population (a 1922 planting at the Federal Experiment Station at Serdang, Malaysia; see Rosenquist, 1986), whereas the second subgroup (DM 714) contains a much lower contribution of 'Serdang Avenue' material in its pedigree.

Greater variation could be generated by introducing material from an unrelated founder population. Thus the material selected from Dami (PNG) and from Binga (DRC) is drawn from the best available in their respective breeding programmes, but the founder populations are only distantly related, if at all. However, intercrossing such material may also lead to loss of breeding progress, given that the populations from each location have been bred for substantially different environments. Nevertheless, the benefits are likely to be greater than crossing within founding populations. Finally, the greatest variation is likely to

come through incorporation of material from wild collections, for example the Tanzanian germplasm analysed here (Fig. 1). Characterizing the amount of variation within such material should allow estimates to be made of how many crosses might be required to capture this additional gene pool efficiently, how different it is from existing material and confirming that it does not represent escapes of previously improved types.

#### *Development and use of the AVROS population*

The above discussion demonstrates the value of markers when broadening populations. However, it is likely that markers will also play a significant role when developing materials within elite populations. To illustrate some of the issues, the AVROS population (represented in most commercial plantings in south-east Asia), will be considered in some detail.

Despite the overwhelming importance of this material, doubt has been expressed as to its authenticity, in particular within the original SP540 selfing where Hartley (1988) points out that of 123 palms in the original population, only two were pisifera, implying that the cross must have been almost entirely incorrect. Supportive evidence for progeny Pol 820 being derived from an outcross comes from comparison of allele variation in a  $F_2$  derived from a recent repeat self-pollination of SP540 (BL605) with conventional AVROS progenies (DM 742). The latter show much greater RFLP band variation, confirming that other unknown palms have contributed to the pedigree (data not shown). Because AVROS material was originally thought to be derived, in the main, from SP540 and hence the Djongo palm, some breeding programmes have avoided intercrossing AVROS palms with African material derived from crosses involving the Djongo palm (e.g. 312/3 and A137/30 in Table 1). However it is clear from Fig. 2 that the latter do not cluster with AVROS material, consistent with the conclusion that AVROS contains errors on the SP540 side at least. Thus errors in pedigree have unnecessarily restricted crossing strategy.

A second application comes through evaluation of the extent of heterozygosity within progenies and whether markers might be used either to maximize residual heterozygosity to avoid inbreeding depression (Hardon, 1970), or indeed to identify individuals which perform well but which have high homozygosity, suggesting the elimination of recessive deleterious genes (Pooni *et al.*, 1989). As an example, the AVROS progeny DM 742 will be considered, because this represents arguably the best pisifera stock currently available — with only four palms contributing pollen for the production of over

12 million seed annually. Of the 157 bands scored, 42 showed segregation within the eight samples of DM 742. This ranged from 17 (40%) heterozygous markers in palm 742.411 to 28 (67%) heterozygous markers in palms DM 742.112 and DM 742.106. Alternatively, similarity indices may be used, and here values ranged from 0.794 (palms DM 742.411 and DM 742.207) to 0.915 (palms DM 742.207 and DM 742.112). Both of these approaches give the breeder additional information to assist in the selection of palms to minimize inbreeding depression or to identify the most uniform progenies.

This study illustrates some of the ways in which DNA markers might be used to assess genetic relationships and assist in the development of breeding programmes. Given that current breeding programmes only involve about 100 parental palms and that selection takes a minimum of nine years, intense genetic analysis is not only feasible but likely to have considerable impact in accelerating breeding progress.

### Acknowledgements

The authors would like to acknowledge the support of Unilever, LonSum and NBPOL plantations which made this work possible, as well as additional samples supplied by RISPA.

### References

- BAUDOIN, L. 1992. Use of molecular markers for oil palm breeding. I. Protein markers. *Oleagineaux*, **47**, 681–691.
- CORLEY, R. H. V. AND LEE, C. H. 1992. The physiological basis for genetic improvement of oil palm in Malaysia. *Euphytica*, **60**, 179–184.
- GHEQUIERE, M. 1984. Enzyme polymorphism in oil palm (*Elaeis guineensis* Jacq.). Genetic control of nine enzyme systems. *Oleagineaux*, **39**, 561–574.
- HARDON, J. J. 1970. Inbreeding in populations of the oil palm (*Elaeis guineensis* Jacq.) and its effect on selection. *Oleagineaux*, **25**, 449–456.
- HARTLEY, C. W. S. 1988. *The Oil Palm*, 3rd edn. Longman Scientific and Technical, Harlow, U.K.
- JACK, P. L., DIMITRIJEVIC, T. A. F. AND MAYES, S. 1995. Assessment of nuclear, mitochondrial and chloroplast markers in oil palm (*Elaeis guineensis* Jacq.). *Theor. Appl. Genet.*, **90**, 643–649.
- KOVACH, W. L. 1993. MVSP — *A multivariate statistical package for IBM PC and compatibles*, version 2.1g. Internet at warrenk@cix.compulink.co.uk
- LEE, C. H., YONG, Y. Y., DONOUGH, C. AND CHIU, S. B. 1990. Selection progress in the Deli dura population. In: *Proceedings of International Workshop on Oil Palm Breeding Populations*, pp. 81–89. Palm Oil Institute of Malaysia, Kuala Lumpur, Malaysia.
- MAYES, S., JACK, P. L., MARSHALL, D. F. AND CORLEY, R. H. V. 1997. The construction of an RFLP linkage map for oil palm. *Genome*, **40**, 116–122.
- MAYES, S., JAMES, C. M., HORNER, S. F., JACK, P. L. AND CORLEY, R. H. V. 1996. The application of restriction fragment length polymorphism for the genetic fingerprinting of oil palm (*E. guineensis* Jacq.). *Mol. Breed.*, **2**, 175–180.
- NEI, M. AND LI, W. H. 1979. Mathematical model for studying genetic variation in terms of restriction endonucleases. *Proc. Natl. Acad. Sci. U.S.A.*, **76**, 5269–5273.
- POONI, H. A., CORNISH, M. A., KEARSEY, M. J. AND LAWRENCE, M. J. 1989. The production of superior lines and second cycle hybrids by inbreeding and selection. *Elaeis*, **1**, 17–30.
- RAJANAIDU, N. 1987. Collection of oil palm (*Elaeis guineensis*) genetic material in Tanzania and Madagascar. *Palm Oil Res. Inst. Malaysia, Bull.*, **15**, 1–6.
- ROSENQUIST, E. A. 1986. The genetic base of oil palm breeding populations. In: *Proceedings of the International Workshop on Oil Palm Germplasm and Utilisation*, pp. 27–56. Palm Oil Institute of Malaysia, Kuala Lumpur, Malaysia.
- SHAH, F. H., RASHID, O., SIMONS, A. J. AND DUNSON, A. 1994. The utility of RAPD markers for the determination of genetic variability in oil palm (*Elaeis guineensis*). *Theor. Appl. Genet.*, **89**, 713–718.
- VIRK, P. S., FORD-LLOYD, B. V., JACKSON, M. T. AND NEWBURY, H. J. 1995. Use of RAPD for the study of diversity within plant germplasm collections. *Heredity*, **74**, 170–179.
- XIAO, J., LI, J., YUAN, L., MCCOUCH, S. R. TANKSLEY, S. D. 1996. Genetic diversity and its relationship to hybrid performance and heterosis in rice as revealed by PCR-based markers. *Theor. Appl. Genet.*, **92**, 637–643.