CHAPTER V

CLUSTERING OF THE MPOB-NIGERIAN OIL PALM GERMPLASM

5.1 INTRODUCTION

Genetic diversity is the variation in the genetic constitution of individuals within or among species and it is the cornerstone of the ability of organisms to adjust to changes in their environment through natural selection. Crop genetic diversity is not only a raw material for industrial agriculture but is of utmost importance to food security and sustainable agriculture as it enables farmers to adapt crops suited to their own site specific ecological needs and cultural traditions (Singh et al., 2008). According to Kumar and Singh (2006), exploitation of heterosis has been the major aim in cross-pollinated crops for which hybrid development or population improvement are commonly adapted procedures. In view of this, genetic variability is requisite to attain genetic benefits in a breeding program.

Characterization of genetic variability and an estimate of the genetic relationship among varieties are crucial to any breeding program. Also, getting accurate estimates of genetic diversity levels among germplasm sources may enhance efficiency of breeding efforts to improve crop species (Singh et al., 2008). Plant breeding deals with high-yielding genotypes and parental selection is the first step in any plant breeding program and this can be done using different selection methods based on the parent performance. Multivariate analysis is one of these methods that can be used in the selection program (Singh et al., 2008).

Cluster analysis is a multivariate statistical technique that has emerged as one of the leading methods of multivariate analysis due to its usefulness (Kettenring, 2006). It was originally developed for biological classification and its main aim in the analysis of data from plant breeding trials is to group the varieties into several homogenous groups such that those varieties within a group have a similar response pattern across the locations (Kroonenberg et al., 1995). The resultant from cluster analysis are usually pictured in the form of hierarchical trees; also called a dendrogram (Siracli et al., 2013). Cluster analysis has been well utilized to classify different crop species. Hayati et al. (2004) made use of cluster analysis to classify 723 accessions of oil palm from 26 populations into three main clusters. Zulkifli et al. (2012) also used cluster analysis in grouping 330 oil palms originating from 11 African countries into three major clusters.

The main goals of this chapter were to use cluster analysis to classify the variation pattern in the MPOB-Nigerian oil palm germplasm, as well identify the germplasm accessions with high genetic diversity to maximize heterosis.

5.2 MATERIALS AND METHODS

5.2.1 Plant Breeding Material

The Malaysian Agricultural Research and Development (MARDI) and the Nigerian Institute for Oil Palm Research (NIFOR) in 1973 jointly made collections of oil palm genetic materials in Nigeria to broaden the genetic base of the oil palm. The Nigerian oil palm germplasm was collected at 45 sites (45 populations) (See Table 1, Figure 1), with an average of 20 palms per site and 200 seeds per palm. A total of 919 (595 *duras* and 324 *teneras*) open-pollinated bunches were harvested during this collection

(Rajanaidu & Rao, 1987). The prospected oil palm breeding materials were planted using eight cubic lattice design on September, 1975 and identified as Trial 0.149 (Rajanaidu & Rao, 1987) at the MPOB Station at Kluang, Malaysia. Kluang is characterized by Rengam soil series and a rainfall of about 2500 mm/yr. However, only the *dura* Nigerian oil palm breeding materials were selected for this research.

5.2.2 Data Collection

Data collections were carried out for bunch yield, bunch quality components, morphological traits, physiological traits and fatty acid composition. The procedures for these traits are presented in **APPENDIX A – D.**

5.2.3 Standardization

Prior to performing cluster analysis on the dataset, data for traits were standardized to a mean of zero and a standard deviation of one to annihilate the biasness due to dissimilar magnitude of the units of measurement (Jolliffe, 2002). The average of the quantitative was standardized using the formula (Microsoft Office Excel, 2010);

$$z = \frac{x - \mu}{\sigma} \tag{7}$$

where;

x = value to standardize

μ = arithmetic mean

σ = standard deviation of the distribution

5.2.4 Cluster Analysis

The standardized data were subjected to two hierarchical clustering analysis methods; single linkage clustering analysis (SLCA) and minimum variance method of WARD (1963) with the aid of "THE UNSCRAMBLER®X" software (CAMO software version 10.1). The theory behind clustering is an expected positive relationship between the variables Euclidean distance and the similarity of the observations. As a result, cluster analysis is driven between minimizing the Euclidean distance of observations within a cluster, and maximizing the Euclidean distance between clusters (Vural & Karasu, 2007). The single linkage (Sneath, 1957) which is an agglomerative hierarchical method merges groups based on the minimum distance between objects in two groups; therefore the distance between Clusters R and Q is defined by:

$$d_s(R,Q) = \min_{i \in R,\, j \in Q} d(i,j) \tag{8}$$

where, $d(i, j)$ is the distance between the i^{th} and j^{th} objects.

The ward's method optimizes an objective function; that is, it minimizes the sum of squares within groups and maximizes the sum of squares between groups. Ward's method is similar to the linkage methods in that it begins with N clusters, each containing one object, it differs in that it does not use cluster distances to group objects. Instead, the total within-cluster sum of squares (SSE) is computed to determine the next two groups merged at each step of the algorithm. The error sum of squares (SSE) is defined (for multivariate data) as:

$$SSE = \sum_{i=1}^{k} \sum_{j=1}^{n_i} (y_{ij} - \bar{y})^2 \tag{9}$$

where y_{ij} is the j^{th} object in the i^{th} Cluster and n_i is the number of objects in the i^{th} Cluster.

Clustering was performed separately based on all the 43 variables; yield, yield components and bunch components; vegetative and physiological traits; as well as fatty acid traits of the MPOB-Nigerian germplasm materials.

5.3 RESULTS

5.3.1 Clustering of MPOB-Nigerian Oil Palm Populations Based on All Variables

5.3.1.1 Single Linkage Clustering (SLCA)

The 44 MPOB-Nigerian oil palm populations based on all the 43 traits were classified into eight groups using single linkage cluster algorithm. The dendrogram resulting from the analysis is presented in Figure 6a. Cluster 1, 2, 3, 4, 5 and 6 were singleton with one population each (2.27 % each) which includes NGA 02, NGA 19, NGA 12, NGA 25, NGA 34 and NGA 1 respectively. Cluster 7 contains 34 populations (77.27 %) namely: NGA 03, 04, 05, 06, 07, 08, 09, 10, 11, 13, 14, 15, 16, 17, 18, 20, 21, 22, 23, 26, 27. 28, 29, 30, 31, 32, 33, 35, 36, 37, 38, 43, 44 and NGA 45. The last cluster, Cluster 8 contains only four populations (9.09 %) namely: NGA 39, NGA 40, NGA 41 and NGA 42.

The mean values of different clusters for all the variables are given in Table 13a. Cluster 1 showed high mean values for MNW, SF, IV and C18:1. Cluster 2 had high mean values for M/F and C18:0. Cluster 3 also showed high mean values for FFB, BNO, MFW, OWM, OB, OY, KY, TEP, LL, LN, BDM, TDM, BI, *e*, *f*, NAR and C16:0. Cluster 4 had high mean values for ABW, FB, PCS, HT and VDM. Cluster 5 also had high mean values for C12:0, C14:0 and C18:2. Cluster 6 showed high mean

values for ODM, RL, LW, LA, LAI, DIAM, LAR and *f* while Cluster 8 had high mean values for K/F, K/B, FP and HT.

5.3.1.2 Ward Hierarchical Clustering (WHCA)

The 44 oil palm populations based 43 variables were classified into eight groups using Ward's hierarchical algorithm as shown in Figure 6b. Cluster 1 contains a single population (2.27 %) and includes NGA 12. Cluster 2 contains two populations (4.55 %) namely: NGA 02 and NGA 19. There were eleven populations in Cluster 3 (25 %) namely: NGA 09, 10, 11, 25, 27, 28, 30, 31, 32, 33, and NGA 37. Cluster 4 contains five populations (11.36 %) namely: NGA 13, 14, 43, 44, and NGA 45. Cluster 5 also contains eleven populations (25 %) which include NGA 05, 06, 07, 08, 20, 21, 22, 23, 35, 36, and NGA 38. Seven populations were in Cluster 6 (15.91 %) namely: NGA 15, 16, 17, 18, 26, 29, and NGA 34. Cluster 7 contains four populations (9.09 %) namely; NGA 39, 40, 41 and NGA 42. Cluster 8 contains three populations (6.82 %) and includes NGA 01, 03 and NGA 04.

The mean values of different clusters for all the variables are given in Table 13b. Cluster 1 showed highest mean values for FFB, BNO, PB, MF, OWM, OB, OY, KY, TEP, LN, BDM, TDM, BI, *e*, NAR, C12:0 and C16:0. Cluster 2 had high mean values for IV, C18:0 and C18:1. Cluster 3 also had high mean values for MFW, MNW and FB. Cluster 4 had high mean value for C16:1. Cluster 5 showed high mean values for ABW, PCS, LL, LW, HT, LA, LAI, VDM and *f*. Cluster 6 also showed high mean value for ODM while Cluster 7 on the other hand showed high mean values for KF, SF, KB, FP, C14:0 and C18:2. The last group, Cluster 8 showed high mean values for RL, DIAM, LAR and *f*.

The least genetic distance was between NGA 08 and NGA 23 (2.272) as shown in the proximity matrix of Euclidean distance (**APPENDIX E1**) while the highest genetic distance was between NGA 12 and NGA 41 (19.326), followed by NGA 12 and NGA 42 (19.302).

5.3.2 Clustering of Oil Palm Populations Based on Yield and Bunch Components

5.3.2.1 Single Linkage Clustering Method (SLCA)

Single linkage Clustering of the MPOB-Nigerian oil palm populations based on the yield, yield components and bunch analysis is shown as a dendrogram in Figure 7a. Four well defined Clusters are visible from the dendrogram. The first (2.27 %) and second Cluster (2.27 %) are singleton with one population each namely NGA 12 and NGA 01 respectively. The third Cluster which is the largest comprised thirty-eight populations (86.36 %) which include NGA 02, 03,04, 05, 06, 07, 08, 09, 11, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 43, 44, and NGA 45. The fourth Cluster comprised four populations (9.09 %) which include NGA 39, 40, 41 and 42. The mean values of different clusters for yield and yield components are given in Table 14a.

Cluster 1 showed high values for FFB, BNO, MFW, PB, MF, OWM, OB, OY, KY and TEP. Cluster 2 had high value for ODM. Cluster 3 on the other hand showed high values for ABW and FB while Cluster 4 showed high values for KF, SF and KB.

5.3.2.2 Ward's Hierarchical Cluster Analysis (WHCA)

Ward's hierarchical cluster analysis based on yield, yield components and bunch components resulted in four visible clusters as shown in the dendrogram (Figure 7b). The first cluster comprised seven populations (15.91 %) which include NGA 01, 02,

03, 04, 43, 44, and NGA 45. The second cluster (61.36 %), which is the largest contains twenty-seven members namely: NGA 05, 06, 07, 08, 09, 10, 11, 13, 17, 18, 20, 21, 22, 23, 25, 26, 27, 28, 29, 30, 31, 32 33, 35, 36, 37, and NGA 38. The third cluster comprised six populations (13.64 %) namely: NGA 12, 14, 15, 16, 19 and NGA 34. The last cluster contains four populations (9.09 %) namely NGA 39, 40, 41, and NGA 42.

The mean values of different clusters for yield and yield components are given in Table 14b. Cluster 2 showed mean high values for ABW, MFW, ODM and FB. Cluster 3 showed high mean values for FFB, PB, MF, OWM, OB, OY, KY and TEP while Cluster 4 showed high mean values for BNO, KF, SF and KB.

As revealed by the proximity matrix of Euclidean distance in **APPENDIX E2**, the least genetic distance was between NGA 29 and NGA 30 (0.961) while the highest genetic distance was between NGA 12 and NGA 42 (13.543), followed by NGA 37 and NGA 42 (13.415).

5.3.3 Clustering of Oil Palm Populations Based on Vegetative and Physiological Traits

5.3.3.1 Single Linkage Clustering (SLCA)

The 44 oil palm populations based on 18 morpho-physiological traits were classified into three clusters using the single linkage clustering method. The dendrogram resulting from the analysis is shown graphically in Figure 8a. Cluster 1 contains a single population (2.27 %); NGA 14. The second cluster had the highest number of populations (86.64 %) with 39 members. The members include NGA 01, 02, 03, 04, 05, 06, 07, 08, 09, 10, 11, 12, 13, 15, 16, 17, 18, 19, 20, 21, 22, 23, 25, 26, 27, 28, 29,

30, 31, 32, 33, 34, 35, 36, 37, 38, 43, 44 and NGA 45. There were four populations in cluster 4 (9.09 %) namely: NGA 39, 40, 41 and NGA 42. The mean values of different clusters for vegetative and physiological traits are given in Table 15a. Cluster 1 showed high mean values for PCS, RL, LL, LW, LN, LA, LAI, DIAM, LAR, BDM, TDM, BI, *e* and *f*. Cluster 2 on the other hand showed high mean values for VDM and NAR while cluster 3 showed high mean values for FP and HT.

5.3.3.2 Ward's Hierarchical Cluster Analysis (WHCA)

The MPOB-Nigerian oil populations based on the 18 vegetative and morphological traits were also classified into three clusters using Ward's method. The dendrogram from the Ward's method is shown in Figure 8b. Cluster 1 had the highest number of members with twenty-eight populations (63.64 %) namely: NGA 01, 02, 03, 04, 05, 06, 07, 08, 09, 10, 11, 14, 20, 21, 22, 23, 25, 27, 28, 29, 30, 31, 32, 33, 35, 36, 37 and NGA 38. Cluster 2 comprised twelve populations (27.27 %) namely: NGA 12, 13, 15, 16, 17, 18, 19, 26, 34, 43, 44 and NGA 45. The last group, cluster 3 had four populations (9.09 %) and they include NGA 39, 40, 41 and 42. The mean values of different clusters for vegetative and physiological traits are given in Table 15b. Cluster 1 showed high mean values for PCS, RL, LL, LW, LN, LA, LAI, DIAM, VDM, TDM, *e* and *f*. Cluster 2 showed high mean values for LN, LAR, BDM, BI and NAR while cluster 3 showed high mean values for FP and HT.

As shown in the proximity matrix of Euclidean distance in **APPENDIX E3**, the least genetic distance based on the 18 vegetative and physiological traits was between NGA 06 and NGA 38 (1.008) while the highest genetic distance was between NGA 14 and 41 (14.554), followed by NGA 20 and NGA 41 (14.544).

5.3.4 Clustering of Oil Palm Populations Based on Fatty Acid Composition

5.3.4.1 Single Linkage Clustering Method (SLCA)

The 44 MPOB-Nigerian oil palm populations were classified into three Clusters based on the oil quality traits. The dendrogram resulting from this analysis is presented in Figure 9a. The first Cluster was a singleton (2.27 %) with just NGA 25 as its member. The second Cluster had the highest number of populations (93.18 %) with forty-one members namely: NGA 01, 03, 04, 05, 06, 07, 08, 09, 10, 11, 12, 13, 14, 15, 16, 17, 18, 20, 21, 22, 23, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44 and NGA 45. The last Cluster had two populations (4.55 %) and includes NGA 02 and NGA 19.

The mean values for different Clusters for fatty acid traits are given in Table 16a. Cluster 2 had high mean values for C12:0, C14:0, C16:0 and C16:1 while Cluster 3 had high mean values for IV, C18:0 and C18:1.

5.3.4.1 Ward's Hierarchical Cluster Analysis (WHCA)

The oil palm populations were also grouped into three clusters using the Ward's clustering method. The resultant dendrogram from the analysis is presented in Figure 9b. Cluster 1 had the highest cluster members with thirty-seven populations (84.09 %) namely: NGA 01, 03, 04, 05, 06, 07, 08, 09, 10, 11, 13, 14, 15, 16, 17, 18, 20, 21, 22, 23, 25, 26, 27, 28, 30, 31, 32, 33, 35, 36, 37, 40, 41, 42, 43, 44 and NGA 45. Cluster 2 comprised five populations (11.36 %) which include NGA 12, 29, 34, 38 and NGA 39. Cluster 3 comprised only two populations (4.55 %) namely: NGA 02 and NGA 19.

The mean values for different clusters for fatty acid traits are given in Table 16b. Cluster 1 showed high mean values for C16:1. Cluster 2 had high mean values for C12:0, C14:0, C16:0 and C18:2 while cluster 3 had high mean values for IV, C18:0 and C18:1.

As revealed in the proximity matrix of Euclidean distance in **APPENDIX E4,** the least genetic distance based on the fatty acid traits was between NGA 16 and NGA 43 (0.686) while the highest genetic distance was between NGA 12 and NGA 19 (11.465), followed by NGA 02 and NGA 12 (11.458).

5.4 DISCUSSION

Samples will be grouped in terms of their nearness or similarity (Hossain et al., 2011). Based on the number of clusters selected during analysis, there is no standard procedure or best criterion to determine the exact number of clusters needed in grouping data (Singh et al., 2008: Hair et al., 1995). The number of clusters selected was based on the PC model used. Furthermore, two clustering methods were used because a single clustering method might not be always optimal and efficient in revealing genetic associations (Mohammadi & Prasanna, 2003). According to Beebe et al. (1998), it is not possible to say what linkage method works best as it depends on the shape of the clusters, therefore, using more than one method is recommended if the aim is to learn as much as possible from the data.

Single linkage clustering analysis (SLCA) was selected from the agglomerative linkage types because it has been reported to be a better technique in that it adequately provides a clearer and more informative display of relative positions of the genotypes (Aremu et al., 2007: Aliyu & Fawole, 2001). Ward's method was also used because it

has generally been considered to be very efficient and the most common approach to doing hierarchical clustering analysis (Siracli et al., 2013: Ferreira & Hitchcock, 2009; Kumar & Singh, 2006).

5.4.1 Clustering Based on All variables

In the clustering of all variables, both SLCA and WHCA differ in some ways in assigning populations to groups. NGA 02 and NGA 19 were placed singly in different clusters as indicated by SLCA while both were placed in the same cluster as shown by Ward's method. These two populations as proved by the PCA results had high scores for iodine value (IV), stearic acid (C18:0) and oleic acid (C18:1). In addition NGA 02 has high percentage of shell-to-fruit content (S/F) as evidenced in the bi-plot of all variables in Figure 3. This may be the reason NGA 02 was in a separate cluster as shown by SLCA. Also, NGA 34, NGA 25 and NGA 01 were also singleton in separate clusters which was not so in Ward's method. Cluster 5 had the highest values for lauric acid (C12:0) and myristic acid (14:0); hence, the reason for being separated out from other members. NGA 25 on the other hand had the lowest value for linoleic acid (C18:1) while NGA 01 has high value for BNO and FB. NGA 12 was both placed in separate cluster by both SLCA and Ward's method; this population had highest value for yield and yield components except in fatty acid content, which is also evident from high PC1 scores of this population. NGA 39, NGA 40, NGA 41 and NGA 42 were placed in the same clusters by both clustering methods and these populations had the lowest values yield and its component and it is also evident from their high scores for traits that are anti-correlated with high yield traits. Hence, when selecting genotypes to be included in the hybridization programme on the basis of

genetic diversity, high cluster means for yield and its components should be given due importance (Kumar & Singh, 2006).

5.4.2 Clustering Based on Yield and Bunch Components

Both SLCA and Ward's method also differ in the assigning of members to groups. In SLCA, NGA 12 and NGA 01 were placed singly in groups while in Ward's method; NGA 12 was placed in the same clusters with NGA 14, 15, 16, 19 and NGA 34 - these populations are high yielding population as evidenced from the result of PCA and also from the same region as indicated in Table 1. NGA 01 on the other hand was placed in the same cluster with NGA 02, 03, 04, 43, 44 and 45. These populations have high values for shell-to-fruit (S/F) and kernel-to-bunch content (K/B) as it is evidenced from the PCA results. From the scores plot in Figure 4, the populations are clustered around each other but NGA 02 was a bit distant from the other members which may be the reason it was placed in singly by SLCA. NGA 39, 40, 41, and NGA 42 were placed together in the same group by both SLCA and Ward's methods. These populations have the lowest values for yield and bunch component variables and also were collected from the same area in Nigeria. Other cluster members have intermediate values for yield and bunch components. It could be concluded that highest mean of FFB, BNO, MFW, MF, OB, OY, KY, TEP from cluster 1 according to SLCA and highest mean values for yield from cluster 2 according to Ward's method could be selected as parents for hybridization program. Ashafaq et al. (2012) also suggested genotypes of rice that could be used for identification of high yielding rice genotypes.

5.4.3 Clustering Based on Vegetative and Physiological Traits

The results of Clustering from both SLCA and Ward's method were also different; NGA 39, 40, 41 and NGA 42 were placed in the same group by both clustering algorithms and members of this group have produced the highest number of fronds (FP). SLCA grouped NGA 14 in a separate group while other populations were put in the same group. This was different in Ward's method as NGA 14 was placed in cluster 1, has members in this group have high values for most of the variables that account for most variation in the vegetative and physiological traits as proven by the PCA results. In addition, cluster 2 which comprised of NGA 12, 13, 15, 16, 17, 18, 19, 26, 34, 43, 44 and 45 as grouped by the Ward's method and members in this group have high values for bunch index (BI) and leaf area ratio (LAR) as proven by the bi-plot of PCA for vegetative and physiological traits. Hence, accessions with lowest height, lowest trunk diameter, highest leaf area, highest leaf area index, BDM, TDM, BI, *e* and *f* from Cluster 1 and Cluster 2 of both SLCA and Ward's method could be selected for hybridization program. Rahman and Mansur (2009) reported similar pattern in lime.

5.4.4 Clustering Based on Fatty Acid Composition

Grouping of members also differs by both methods, but in both methods NGA 02 and NGA 19 were placed in the same clusters as these two populations have the highest value for iodine value, stearic acid (C18:0) and oleic acid (C18:1). These populations can be selected in breeding for these particular oil quality traits. Also NGA 25 was placed singly in a cluster by SLCA because this population had the lowest value for linoleic acid (C18:2). Grouping by Ward's method was different, Cluster 1 which had the highest number of members comprised of populations almost same values for fatty

acids traits, while Cluster 2 comprised of NGA 12, 29, 34, 38 and NGA 39. From the PCA biplot for fatty acid traits in Figure 5e, these populations can be seen close to one another and members in this cluster have high values for lauric acid (C12:0) and myristic acid (C14:0). Hence, for breeding for good oil quality palms accessions with high mean values for fatty acid traits can be selected. Khan et al. (2009) found similar trend in safflower.

Of the two clustering methods, Ward's method was more in line with the results from the PCA than single linkage clustering method. Many researchers have confirmed the efficacy of Ward's method in clustering. Blashfield (1976) compared four types of hierarchical clustering methods (single linkage, complete linkage, average linkage and Ward's method) for accuracy in recovery of original population clusters. Cohen's statistic was used to measure the accuracy of the clustering methods and Ward's method performed significantly better than the other clustering procedures. Hands and Everitt (1987) also compared five hierarchical clustering techniques (single linkage, complete linkage, average linkage, centroid and Ward's method) on multivariate binary data and discovered that Ward's method was the best overall than other hierarchical methods. Ferreira and Hitchcock (2009) also confirmed that Ward's method was usually the best, likewise Milligan and Cooper (1980) who found that single linkage technique was least effective while Ward's methods gave the best overall recovery. Kumar and Singh (2006) in addition, identified Ward's method as the best clustering solution for grouping of the inbred lines.

5.4.5 Genetic Distance

Mohammadi and Prasanna (2003) stated that analysis of genetic relations in crop species is a significant component of crop species, as it serves to provide information

about genetic diversity and also a platform for stratified sampling of breeding populations. Genetic diversity can be analyzed at different in crop plants; individual genotypes such as inbred lines, pure lines or clone, populations, germplasm accessions, and species (Mohammadi & Prasanna, 2003). Genetic distance according to Nei (1973) "is that difference between two entities that can be described by allelic variation". This was further detailed by him (1987) as "the extent of gene differences between populations or species that is measured by some numerical quantity". Euclidean or straight line measure of distance is the most commonly used statistical method for calculating distance between individuals (genotypes or populations) by morphological data (Ferreira & Hitchcock, 2009; Mohammadi & Prasanna, 2003).

The proximity matrix of Euclidean distance for all the variables studied showed that the populations NGA 12 and NGA 41 had the highest genetic distance and were grouped in different clusters followed by NGA 12 and NGA 42 which were also in different clusters. For the yield and bunch quality components, highest genetic distance was between NGA 12 and NGA 42 followed by NGA 37 and NGA 42 while the least was between NGA 29 and NGA 30. The highest genetic distance for the vegetative and physiological traits was between NGA 14 and NGA 41 and was followed by NGA 20 and NGA 41. The least genetic distance for these traits was encountered between NGA 06 and NGA 38. Also, for the fatty acid traits, highest genetic distance was between NGA 12 and NGA 19 followed by NGA 02 and NGA 12 while the least genetic distance was between NGA 16 and NGA 43.

From this study, populations with highest genetic distance between them were in different clusters while those with the least genetic distances belong to the same clusters. In view of this, it can be anticipated that crossing between morphologically

distant populations will result to maximum degree of heterosis (Odewale et al., 2012; Singh et al., 2008; Mohammadi & Prasanna, 2003). Hence, the use of these populations for breeding programs should be given utmost importance for maximum heterosis as it is expected to produce new recombinants with desired characters and high hybrid vigor. On the other hand, crossing between populations with the least genetic distance between them should be avoided (Odewale et al., 2012a). Odewale et al. (2012b) observed similar result in coconut germplasm. However, it was stated by Rahim et al. (2010) that genotypes with minimum distance between them could be used for backcross breeding program.

Furthermore, hybridization between populations of different clusters with high cluster mean will result into palms which will perform better than their parents (Kumar et al., 2010). Therefore, populations from the clustering of all variables in Clusters 1 which include NGA 12 and Clusters 6 which includes NGA 15, 16, 17, 18, 26, 29, and NGA 34 as grouped by Ward's method, which showed high scores for important characters like yield, yield components and oil quality traits can be used in future breeding programme to recombine these characters. Kumar et al. (2010) similar results in advanced breeding lines of groundnut,

According to Rahman and Al-mansur (2009), higher inter and intra cluster distances indicate higher genetic variability between and within clusters respectively and the minimum inter and intra cluster distances indicate closeness among the accessions of two clusters and within the clusters. The high level of genetic variability encountered in this MPOB-Nigerian oil palm germplasm was also confirmed in the research of Zulkifli et al. (2012). They found out that this Nigerian germplasm showed the highest

number of alleles per locus and had the highest number of rare alleles, a high percentage of polymorphic loci and high heterozygosity.

Genetic diversity is usually associated with geographical diversity but genetic diversity is not necessarily directly related with geographical distribution (Rahman & Al-Mansur, 2009). In this study, grouping of the populations by the two clustering did not follow a particular group. Some populations from same region were grouped together while others from different areas were clustered together. This suggests that there is no correspondence between population and their geographic origin. Though Kjaer et al. (2004) found correlation between genetic distance and geographic locations in sago palm (a member of the oil palm family; Arecaceae). Zulkifli et al. (2012) also found a strong association between genetic distance and geographic location in evaluation of MPOB oil palm germplasm populations. However, the absence of association between genetic distance and geographic location in this study suggests that the populations of different locations have genetic similarity and could have been derived from the same breeding materials (Tahir et al., 2013) or the random distribution of populations into various clusters from different geographic location implicates that drives other than geographic influence such as exchange of breeding material, genetic drift, natural and artificial selections are responsible for diversity as reported by Murthy and Arunachalm (1966). These findings also follow the same pattern with studies on crops like groundnut (Kumar et al. 2010; Makinde & Ariyo, 2010), sugarcane (Tahir et al., 2013), lime (Rahman & Al-Mansur, 2009), sesame (Seymus & Uzun, 2010), safflower (Khan et al., 2009), coconut (Odewale et al., 2012), soybean (Iqbal et al., 2008) and other crops.

5.5 CONCLUSION

Cluster analysis as a multivariate analysis has proven as a useful tools in a number of ways. Firstly, it has helped to group the oil palm accessions into groups based on similarity. Secondly, it has helped to identify populations that can best be combined for specific traits. Lastly, populations that are genetically distant from each other have been identified and these can be used for selection of oil palms with hybrid vigor. Populations with good amount of genetic divergent (NGA 12 and NGA 41; NGA 12 and NGA 02; NGA 12 and NGA 19) and superior agronomic characters (NGA 12, NGA 14, NGA 15, NGA 16, NGA 19 and NGA 34) will be of practical use and can be used to derive superior offspring. The results from the study also showed no consistency of genetic distance with geographic location as populations from different region were similar and populations from same region differed in traits. This diversity could serve as a source of elite materials for oil palm improvement.

Figure 6: Dendrograms of 44 MPOB-Nigerian oil palm germplasm based on all 43 characters of bunch (a) Single linkage Cluster method; (b) Ward Clustering method

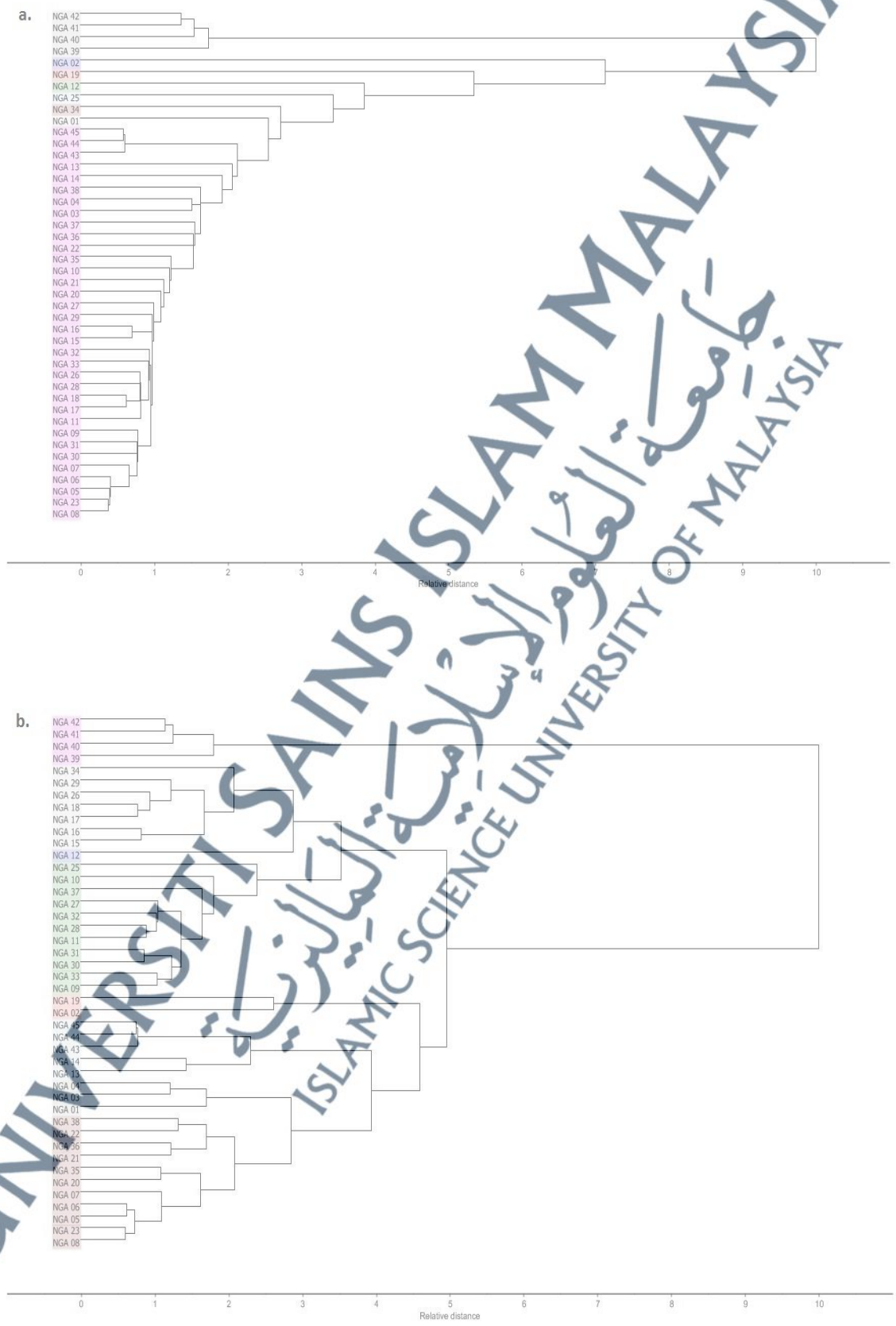

Table 13a: SLCA Cluster means of all variables in MPOB-Nigerian oil palm germplasm

Traits	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5	Cluster 6	Cluster 7	Cluster 8
FFB	136.62	153.07	**190.29**	151.94	153.64	145.25	156.80	102.07
BNO	13.79	12.57	**17.23**	12.71	13.72	15.34	13.18	15.33
ABW	10.29	12.48	11.22	**12.81**	11.65	9.84	12.45	7.02
MFW	8.91	7.47	**9.69**	9.63	8.68	8.77	9.67	6.68
MNW	**5.29**	3.78	4.96	5.11	4.62	4.94	5.24	4.08
PB	1.96	3.45	**5.08**	2.32	2.76	2.55	2.39	2.58
MF	40.71	**49.33**	48.55	46.97	46.72	43.62	45.79	38.86
KF	12.81	11.67	12.24	11.97	13.16	12.10	12.31	**15.58**
SF	**46.48**	39.00	39.22	41.06	40.12	44.27	41.90	45.56
ODM	72.32	73.83	75.07	74.86	74.24	**75.28**	74.63	73.26
OWM	45.14	48.58	**50.57**	49.26	48.51	50.05	48.78	48.56
FB	65.34	66.03	65.44	**67.08**	65.70	63.68	66.99	66.42
OB	12.08	15.80	**16.13**	15.54	14.94	13.87	14.99	12.53
KB	8.10	7.29	7.41	7.74	8.25	7.41	7.95	**9.93**
OY	16.43	23.86	**30.21**	23.71	23.11	20.24	23.59	12.89
KY	10.96	10.97	**13.86**	11.81	12.70	10.91	12.45	10.22
TEP	23.01	30.44	**38.52**	30.80	30.73	26.79	31.06	19.03
FP	28.20	25.98	25.64	28.36	27.09	28.42	27.42	**30.67**
PCS	14.47	14.96	15.69	**16.58**	14.39	15.83	16.25	10.39
RL	4.34	4.67	4.67	4.53	4.53	**4.76**	4.67	3.57
LL	79.43	84.25	**85.88**	82.87	81.71	83.46	84.99	68.86
LW	4.31	4.36	4.24	4.44	4.15	**4.56**	4.40	3.85
LN	154.86	157.34	**162.64**	158.94	154.67	160.38	159.41	147.71
HT	1.05	1.10	1.02	**1.33**	0.98	1.04	1.27	**1.33**
LA	6.08	6.69	6.78	6.71	5.98	**6.96**	6.85	4.52
LAI	3.60	3.96	4.02	3.97	3.54	**4.12**	4.05	2.68
DIAM	0.71	0.66	0.67	0.70	0.64	**0.75**	0.68	0.64
LAR	20.62	22.21	21.86	19.97	21.25	**22.22**	20.95	19.65
BDM	10.66	12.42	**15.22**	12.07	12.05	11.39	12.42	8.10
VDM	8.35	7.81	7.96	**9.61**	7.73	9.09	9.04	7.10
TDM	19.01	20.23	**23.18**	21.67	19.78	20.48	21.46	15.20
BI	0.56	0.61	**0.65**	0.56	0.60	0.56	0.58	0.53
E	0.79	0.81	**0.92**	0.86	0.82	0.80	0.85	0.75
F	0.77	0.80	**0.82**	0.81	0.77	**0.82**	0.81	0.66
NAR	10.35	10.06	**11.27**	10.71	10.86	9.69	10.41	11.18
IV	**57.49**	57.15	50.82	54.30	54.59	54.70	54.15	54.14
C12:0	0.00	0.00	1.00	0.73	**2.10**	0.01	0.77	0.99
C14:0	0.40	0.40	1.35	1.05	**1.80**	1.02	1.22	1.39
C16:0	33.10	32.80	**42.35**	38.24	36.80	39.42	38.85	38.79
C16:1	0.00	0.00	0.00	0.01	0.00	**0.10**	0.07	0.03
C18:0	8.60	**9.60**	5.75	6.16	6.90	6.58	6.23	5.94
C18:1	**46.90**	45.30	38.55	42.65	39.10	41.00	41.20	41.09
C18:2	9.90	10.50	10.20	0.00	**12.10**	10.97	10.57	10.69

* Figures in bold indicates maximum value

Table 13b: Ward's Cluster means of all variables in MPOB-Nigerian oil palm germplasm

Traits	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5	Cluster 6	Cluster 7	Cluster 8
FFB	**190.29**	144.85	157.06	160.31	151.50	159.63	102.07	145.86
BNO	**17.23**	13.18	13.19	13.98	12.08	13.64	15.33	14.32
ABW	11.22	11.39	12.45	11.98	**13.15**	12.16	7.02	10.62
MFW	9.69	8.19	**10.24**	9.02	9.71	9.26	6.68	9.00
MNW	4.96	4.54	**5.46**	5.03	5.37	4.82	4.08	5.16
PB	**5.08**	2.71	2.26	2.24	2.22	2.81	2.58	1.82
MF	**48.55**	45.02	46.72	44.27	44.74	47.96	38.86	42.65
KF	12.24	12.24	11.61	13.90	12.50	12.08	**15.58**	12.10
SF	39.22	42.74	41.68	41.83	42.76	39.96	**45.56**	45.25
ODM	75.07	73.08	75.22	73.51	74.39	**75.02**	73.26	74.24
OWM	**50.57**	46.86	49.15	47.50	48.63	49.17	48.56	48.69
FB	65.44	65.69	**67.48**	67.14	66.82	66.69	66.42	65.45
OB	**16.13**	13.94	15.52	14.13	14.54	15.75	12.53	13.57
KB	7.41	7.69	7.57	9.02	8.06	7.71	**9.93**	7.71
OY	**30.21**	20.15	24.48	22.70	22.04	25.37	12.89	19.84
KY	**13.86**	10.96	11.90	14.39	12.19	12.35	10.22	11.36
TEP	**38.52**	26.73	31.63	31.33	29.36	32.78	19.03	26.66
FP	25.64	27.09	28.26	25.67	27.69	27.29	**30.67**	27.76
PCS	15.69	14.71	16.33	15.78	**17.29**	15.04	10.39	15.91
RL	4.67	4.51	4.60	4.74	4.70	4.58	3.57	**4.78**
LL	85.88	81.84	85.14	82.52	**86.50**	82.85	68.86	85.15
LW	4.24	4.34	4.35	4.43	**4.47**	4.29	3.85	4.63
LN	**162.64**	156.10	157.94	160.40	160.72	157.82	147.71	159.36
HT	1.02	1.07	1.35	1.02	**1.38**	1.16	1.33	1.18
LA	6.78	6.38	6.72	6.73	**7.13**	6.43	4.52	7.18
LAI	4.02	3.78	3.98	3.98	**4.22**	3.81	2.68	4.25
DIAM	0.67	0.69	0.68	0.68	0.70	0.66	0.64	**0.73**
LAR	21.86	21.42	20.40	21.50	20.42	21.34	19.65	**22.32**
BDM	**15.22**	11.54	12.42	12.66	11.94	12.67	8.10	11.70
VDM	7.96	8.08	9.38	8.07	**9.75**	8.27	7.10	9.04
TDM	**23.18**	19.62	21.80	20.73	21.69	20.94	15.20	20.74
BI	**0.65**	0.58	0.57	0.61	0.55	0.60	0.53	0.56
E	**0.92**	0.80	0.87	0.82	0.84	0.85	0.75	0.80
F	0.82	0.79	0.81	0.81	**0.83**	0.79	0.66	**0.83**
NAR	**11.27**	10.21	10.76	10.22	10.11	10.78	11.18	9.58
IV	50.82	**57.32**	54.33	53.72	54.62	53.97	54.14	54.50
C12:0	**1.00**	0.00	0.83	0.80	0.68	0.94	0.99	0.41
C14:0	1.35	0.40	1.19	1.27	1.17	1.30	**1.39**	1.14
C16:0	**42.35**	32.95	38.28	39.29	38.51	38.94	38.79	39.15
C16:1	0.00	0.00	0.05	**0.09**	0.08	0.05	0.03	0.11
C18:0	5.75	**9.10**	6.34	6.39	6.18	6.26	5.94	6.14
C18:1	38.55	**46.10**	41.72	40.33	41.71	40.58	41.09	41.34
C18:2	10.20	10.20	9.56	10.68	10.57	10.83	**10.69**	10.64

* Figures in bold indicates maximum value

Figure 7: Dendrograms of 44 MPOB-Nigerian oil palm germplasm based on 17 yield and bunch quality traits (a) Single linkage Cluster analysis; (b) Ward Clustering method

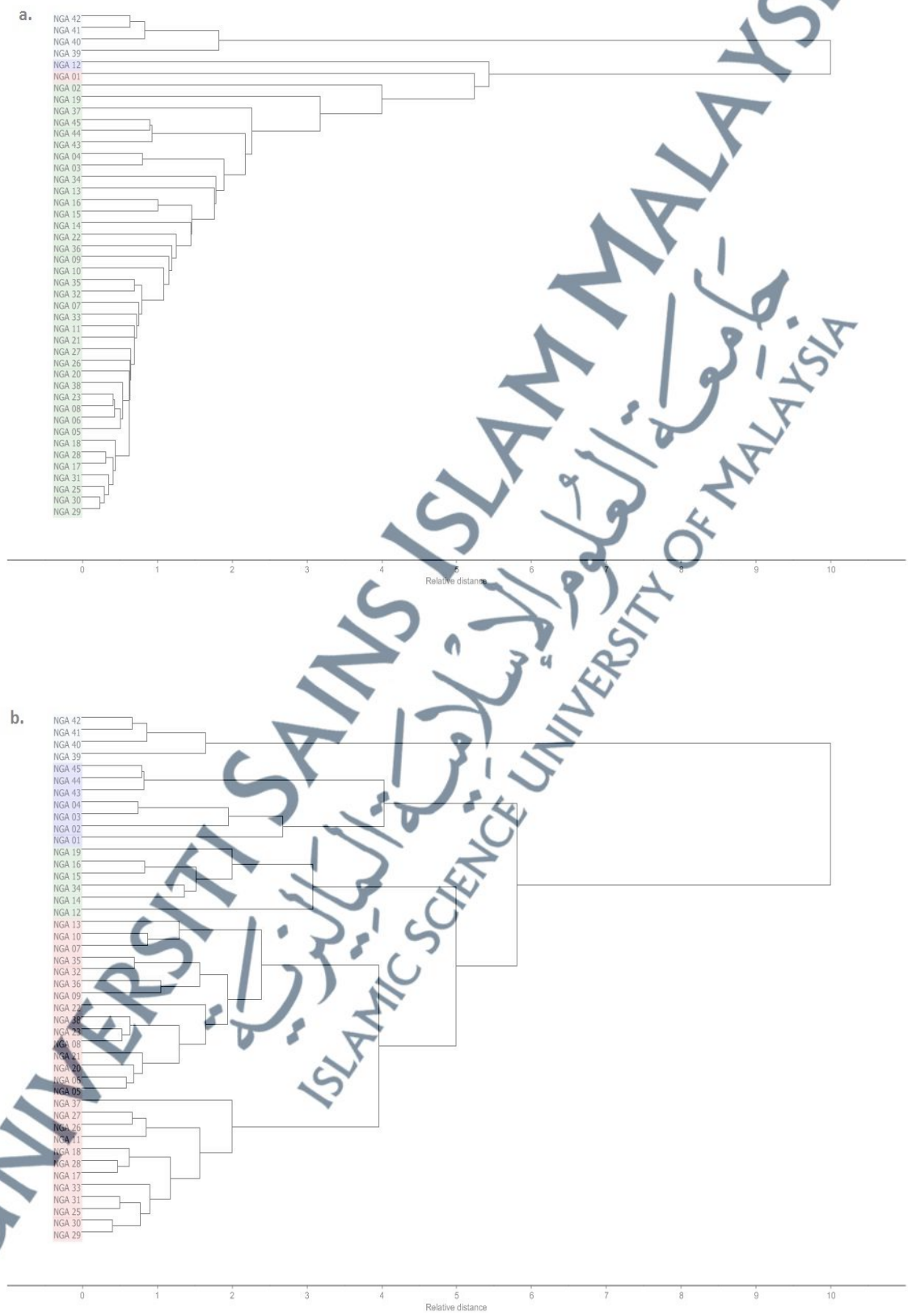

Table 14a: SLCA Cluster means of yield and yield components in MPOB-Nigerian oil palm germplasm

Traits	Cluster 1	Cluster 2	Cluster 3	Cluster 4
FFB	**190.29**	145.25	155.14	102.07
BNO	**17.23**	15.34	13.09	15.33
ABW	11.22	9.84	**12.41**	7.02
MFW	**9.69**	8.77	9.58	6.68
MNW	4.96	4.94	**5.20**	4.08
PB	**5.08**	2.55	2.33	2.58
MF	**48.55**	43.62	45.72	38.86
KF	12.24	12.10	12.31	**15.58**
SF	39.22	44.27	41.96	**45.56**
ODM	75.07	**75.28**	74.53	73.26
OWM	**50.57**	50.05	48.61	48.56
FB	65.44	63.68	**66.94**	66.42
OB	**16.13**	13.87	14.90	12.53
KB	7.41	7.41	7.95	**9.93**
OY	**30.21**	20.24	23.22	12.89
KY	**13.86**	10.91	12.33	10.22
TEP	**38.52**	26.79	30.62	19.03

* Figures in bold indicates maximum value

Table 14b: Ward's Cluster means of yield and yield components in MPOB-Nigerian oil palm germplasm

Traits	Cluster 1	Cluster 2	Cluster 3	Cluster 4
FFB	147.86	155.05	**167.83**	102.07
BNO	13.83	12.86	14.37	**15.33**
ABW	11.17	**12.62**	12.12	7.02
MFW	8.91	**9.82**	8.80	6.68
MNW	5.13	**5.29**	4.54	4.08
PB	2.00	2.34	**3.32**	2.58
MF	42.47	46.15	**48.44**	38.86
KF	13.33	12.01	12.33	**15.58**
SF	44.19	41.85	39.24	**45.56**
ODM	73.48	**74.77**	74.70	73.26
OWM	47.56	48.88	**49.04**	48.56
FB	66.18	**67.13**	66.00	66.42
OB	13.37	15.16	**15.70**	12.53
KB	8.57	7.77	7.72	**9.93**
OY	19.90	23.59	**26.28**	12.89
KY	12.74	12.05	**12.87**	10.22
TEP	27.54	30.82	**34.00**	19.03

* Figures in bold indicates maximum value

Figure 8: Dendrograms of 44 MPOB-Nigerian oil palm germplasm based on 18 vegetative and physiological traits (a) Single linkage Cluster analysis; (b) Ward Clustering method

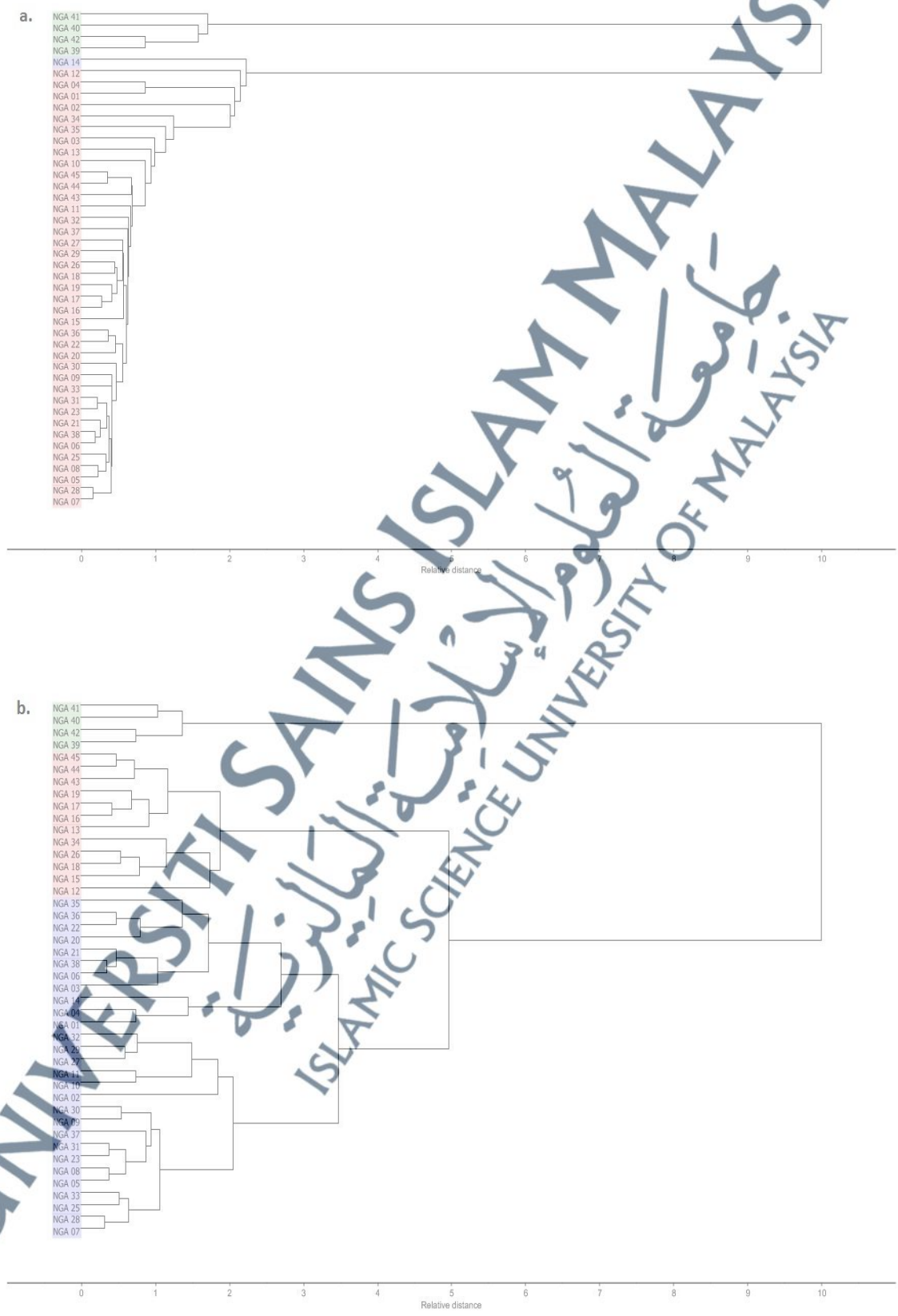

Table 15a: SLCA Cluster means of vegetative and physiological traits in MPOB-Nigerian oil palm germplasm

Traits	Cluster 1	Cluster 2	Cluster 3
FP	26.83	27.47	**30.67**
PCS	**16.50**	16.16	10.39
RL	**4.88**	4.65	3.57
LL	**87.76**	84.56	68.86
LW	**4.55**	4.39	3.85
LN	**162.83**	159.03	147.71
HT	1.05	1.25	**1.33**
LA	**7.46**	6.78	4.52
LAI	**4.42**	4.02	2.68
DIAM	**0.70**	0.69	0.64
LAR	**22.81**	20.89	19.65
BDM	**13.66**	12.30	8.10
VDM	8.79	**9.00**	7.10
TDM	**22.46**	21.30	15.20
BI	**0.61**	0.57	0.53
E	**0.86**	0.85	0.75
F	**0.84**	0.81	0.66
NAR	10.06	**10.42**	11.18

* Figures in bold indicates maximum value

Table 15b: Ward's Cluster means of vegetative and physiological traits in MPOB-Nigerian oil palm germplasm

Traits	Cluster 1	Cluster 2	Cluster 3
FP	27.93	26.34	**30.67**
PCS	**16.55**	15.30	10.39
RL	**4.66**	4.65	3.57
LL	**85.52**	82.60	68.86
LW	**4.43**	4.33	3.85
LN	**159.12**	**159.12**	147.71
HT	1.32	1.08	**1.33**
LA	**6.92**	6.53	4.52
LAI	**4.09**	3.87	2.68
DIAM	**0.70**	0.66	0.64
LAR	20.72	**21.45**	19.65
BDM	12.18	**12.82**	8.10
VDM	**9.40**	8.05	7.10
TDM	**21.53**	20.87	15.20
BI	0.56	**0.61**	0.53
E	**0.85**	0.84	0.75
F	**0.82**	0.80	0.66
NAR	10.34	**10.58**	11.18

* Figures in bold indicates maximum value

Figure 9: Dendrograms of 44 MPOB-Nigerian oil palm germplasm based on eight fatty acid traits (a) Single linkage Cluster analysis; (b) Ward Clustering method

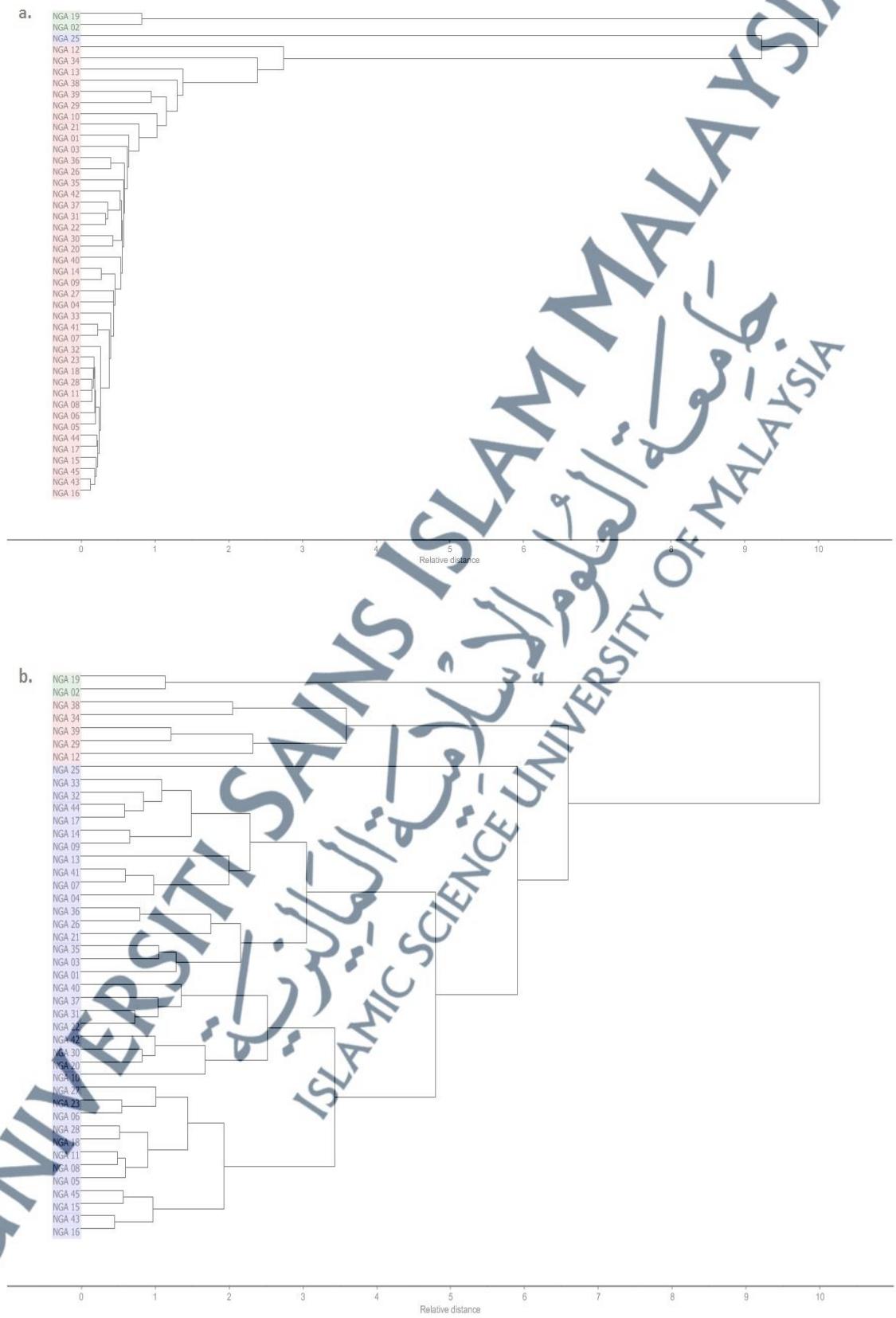

Table 16a: SLCA Cluster means of fatty acid traits in MPOB-Nigerian oil palm germplasm

Traits	Cluster 1	Cluster 2	Cluster 3
IV	54.30	54.18	**57.32**
C12:0	0.73	**0.80**	0.00
C14:0	1.05	**1.24**	0.40
C16:0	38.24	**38.79**	32.95
C16:1	0.01	**0.06**	0.00
C18:0	6.16	6.23	**9.10**
C18:1	42.65	41.16	**46.10**
C18:2	0.00	**10.63**	10.20

* Figures in bold indicates maximum value

Table 16b: Ward's Cluster means of fatty acid traits in MPOB-Nigerian oil palm germplasm

Traits	Cluster 1	Cluster 2	Cluster 3
IV	54.34	53.01	**57.32**
C12:0	0.70	**1.54**	0.00
C14:0	1.18	**1.64**	0.40
C16:0	38.69	**39.45**	32.95
C16:1	**0.07**	0.03	0.00
C18:0	6.26	5.95	**9.10**
C18:1	41.47	39.24	**46.10**
C18:2	10.28	**11.03**	10.20

* Figures in bold indicates maximum value