MJoSHT
Malaysian Journal of Science, Health & Technology
[mjosht.usim.edu.my]

USIM

Article

Evaluation of Oil Palm Germplasm from Senegal and Gambia Using Chemometric Techniques

Saleh H.[1], Li-Hammed M.A[1], Kushairi A.[3], Rajanaidu N.[3], Mohd Sukri H.[1], Che Wan Ngah Z.[1,2], Jalani, B.S.[1]

[1]*Faculty of Science and Technology, Universiti Sains Islam Malaysia (USIM), 71800 Nilai, Negeri Sembilan, Malaysia*
E-mail: cwzanariah@gmail.com

[2]*Institute of Halal Research and Management (IHRAM), Universiti Sains Islam Malaysia (USIM), 71800 Nilai, Negeri Sembilan, Malaysia*
[3]*Malaysian Palm Oil Board (MPOB), P.O. Box 10620, 50720 Kuala Lumpur, Malaysia*

Abstract-- Estimation of genetic diversity and determination of the relationships between collections are useful strategies for ensuring efficient germplasm collection and utilization. Oil palm germplasm materials collected from Senegal and Gambia maintained at the Malaysian Palm Oil Board (MPOB) Kluang Station were characterized for genetic diversity. A total of 44 agronomic traits of these oil palm materials was subjected to simple statistics to evaluate the genetic variability; and to chemometric techniques (Principal Component Analysis and Cluster analysis) to identify the characters contributing to the overall variation and classify the materials based on similarity. The results of the variability profile showed that the Senegal and Gambia oil palm germplasm exhibited low to high variability for the various traits. Nine principal components with eigenvalue >1 accounted for 88 % of the total variation with Principal Component 1 capturing majority of the variation. Most of the traits especially oil yield and yield component traits, contributed to the divergence between and within the germplasm, indicating that wide variation exists in the germplasm materials studied. Ward's cluster analysis based on the PCA results grouped the 42 oil palm accessions into six clusters with cluster-VI having the highest number of members. Furthermore, there was no association between genetic diversity and geographical origin. The means of the agronomic traits of each cluster showed that cluster-III had the highest mean value of yield traits. Also, the cluster groups having high mean values for desired traits could be selected for the traits per se. These accessions could be used to produce high yielding oil palm materials.

Keywords: **Oil palm; genetic diversity; chemometrics; principal component analysis; cluster analysis.**

I. INTRODUCTION

Oil palm (*Elaies guineensis* Jacq.) is a diploid monocotyledon belonging to the family Arecaceae. The oil derived from it, palm oil is one of the world's most traded vegetable oil in the international market and most widely consumed edible oil [1], [2]. It has been predicted that by the year 2020, the world production of oils and fats will increase to 174 million tonnes and palm oil production to 35 million tonnes and that by then, palm oil will be the dominant vegetable oil in the world [3]. Most of oil palm plantation and palm oil production are provided by Indonesia and Malaysia, as both contribute about 44 % and 41.5 % respectively to the world palm oil production [4]. However, the oil palm breeding populations in both countries are derived from a narrow genetic pool. Most of the commercial planting materials utilized are derived from the Deli *dura*, which was first introduced in Indonesia and afterwards in Malaysia [1]. The narrow genetic pool of oil palm resulted into quite a lot of expeditions being mounted by researchers in Malaysia to collect oil palm germplasm in Africa and south-central America [5].

Evaluation of oil palm genetic variability is needed for various purposes such as for the selection of superior palms and it also serves as a first step towards effective utility in breeding programs [6]. With the dawn of advanced computer technologies, it has become possible to study the complex relationship among genotypes through chemometric or multivariate analyses which offers an enhanced understanding of the structure, predominantly of large germplasm collections [7]. In this view, chemometric or multivariate statistical methods, particularly the principal component and cluster analyses have gained wide recognition in the evaluation of germplasm materials of many species [8], [9], [10], [11], [12], [13], [14].

Chemometric is the science of relating measurements made on a chemical system or process for the complex state of the system via multivariate statistical methods [15]. One of the

main advantages of chemometric methods is that it is possible to explore complex co-linear multivariate information in a graphic display. Principal component analysis (PCA) is the fundamental chemometric method based on vector algebra [15]. The main purpose of the method is to reduce the dimensions of complex multivariate data and to simplify data interpretation by finding new orthogonal variables, principal components (PCs), describing the variance in the data. Cluster analysis is used in the categorization of germplasm materials into groups based on similarity or dissimilarity [2].

There is a need to introduce oil palm germplasm as the genetic base of the oil palm industry in Malaysia is very narrow, mostly originating from the four palms planted in Bogor in 1848. Hence, the germplasm introduced needs to be evaluated and characterized. This study, therefore, aims to determine the level of variation in the oil palm germplasm and to identify and classify the groups of accessions with different genetic diversity based on quantitative traits.

II MATERIALS AND METHODS

A. Breeding Materials and Site Location

The germplasm used in this study originated from Senegal and Gambia. A random sample of five families from the breeding materials was collected in July-August 1993 by researchers from Malaysian Palm Oil Board (MPOB) from the two countries. The germplasm materials were planted at Kluang under the MPOB Research Station Kluang, Johor in 1996. The palms were derived from the Independent Completely Randomized Design, where a total number of 415 open-pollinated palms were planted in Trial 0.352 (Senegal materials) and Trial 0.357 (Gambia materials) with two replicates each (41 progenies in replication 1 and 15 progenies in replication 2). For the purpose of this study, available quantitative data on the Senegal and Gambia germplasm were collected at the MPOB headquarter.

B. Data Collection

Data on yield and yield components were evaluated from 2000 – 2007. Harvesting of oil palm usually begins at 36 months after field planting with subsequent operations carried out at regular intervals of seven to ten days, i.e. three rounds in a month. The bunch yield components were fresh fruit bunch (FFB), bunch number (BNO), and average bunch weight (ABW) and other derivatives used in this study. Quantitative data on the bunch components were evaluated in the 2001 until 2006. Bunch and fruit components were determined using the bunch analysis technique developed by Blaak *et al*. [16]. Vegetative characters of the oil palm germplasm were pooled eight years after field planting. Frond production was first calculated after the 7th year before other parameters were taken a year after i.e. year 2004. The physiological characters on the other hand were assessed based on measurements on collective data of bunch yields and bunch quality components, following the methods developed by Squire [17]. Data on the physiological characters were assessed in the year 2007. Fatty acid traits on the other hand were pooled between the years 2000 – 2007. The fatty acid composition was evaluated using the method proposed by Timms [18] for routine analysis using Gas chromatography.

C. Statistical Analysis

The quantitative morphological data collected was arranged in Excel Microsoft word. The oil palm accessions were organized according to their family codes. Simple descriptive statistics such as mean, standard deviation, standard error, minimum, maximum and variance for each collected trait were calculated using SPSS statistical tool. This was done to know the extent of variation in the germplasm accessions. The average of the quantitative data was also standardized to give equal weight to all measurements using the following formula:

$$Z = X - \mu/\sigma \quad (1)$$

where, X = Value to standardize

μ = Arithmetic mean

σ = standard deviation of the distribution

D. Principal Component Analysis (PCA)

PCA simplifies the complex data by transforming a number of correlated variables into a smaller number of variables called principal components. The first principal component accounts for maximum variability in the data as compared to each succeeding component. PCA was analyzed using "THE UNSCRAMBLER®X" software [19]. Mathematically, PCA involved in the decomposition of the original data matrix, X, into a structure part and noise part. In matrix representation, the model with a given number of components as follows:

$$X = TP^T + E \quad (2)$$

where T is the scores matrix, P the loadings matrix (transposed) and E the error matrix. The structured part of the data are the combination of scores and loadings in which focused by user on interpretation of PCA results while the remaining part is called error or residual matrix. The a^{th} column of T and a^{th} row of P is represented by vectors of t_a and p_a respectively, and both are the vector representations of the a^{th} PC. The number of PCs is denoted by A, while a is the number of PC such as 1, 2, 3 up to A. The maximum number of PCs (A) determined is either $I - 1$ (number of objects – 1) or J (number of variables) depending on which give smaller value. Thus, the first scores vector and the first loadings vector are called eigenvectors of the first principal component. Therefore, each successive component is characterized by a pair of eigenvectors for both the scores and loadings [19].

E. Cluster Analysis

Cluster analysis identifies variable which were further clustered into main groups and subgroups using Ward's method through the "THE UNSCRAMBLER®X" software [19]. The ward's method [20] used in this study 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:

$$SSE = \sum_{i=1}^{k} \sum_{j=1}^{n_i} (y_{ij} - \bar{y})^2 \quad (3)$$

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.

III. RESULTS AND DISCUSSION

A. Principal Component Analysis

It was stated that when agronomic character contributed mostly to yield, it complicates the designing of an ideal crop architecture and that finding out several typical agronomic traits of a crop is of assistance in architecture designing and / or in designing high crop yield. This concept could be materialized through multivariate statistical analysis [10]. PCA through its dimension reduction method is of immense help in knowing the traits contributing most to variation.

In present study, nine PCs with Eigen values greater than one and total cumulative variance of 88 % were extracted from the numerous variables through PCA with average bunch weight (ABW) and bunch weight (BWT), petiole cross section (PCS), bunch number (BNO), palmitic acid (C16:0), carotene, mean nut weight (MNW), plant height (HT), iodine value (IV) and kernel to bunch ratio (K/B) contributing mostly to PC1, PC 2, PC 3, PC 4, PC 5, PC 6, PC 7, PC 8 and PC 9, respectively (Table 1). This translates that oil palm accession with higher scores for these traits seems will attain high yield more easily and as a result, these traits should be given utmost importance in the breeding program of oil palm germplasm under study. This result follows similar trends to that of Deyong [10] in analysis among main agronomic traits of spring wheat. Furthermore, it can also be observed from the PCA results that most of the yield contributing traits were poor on other PCs except for PC 1. From the findings of this study, it is obvious that a good hybridization breeding program can be initiated by the selection of genotypes from PC 1 and PC 2. This result also agrees with the findings on morphological diversity and trait association in bread wheat [21].

TABLE I

PRINCIPAL COMPONENT ANALYSIS FOR SENEGAL AND GAMBIA OIL PALM GERMPLASM BASED ON 46 AGRONOMIC TRAITS

	Principal Component Axes								
Traits	1	2	3	4	5	6	7	8	9
FFB (fresh fruit bunch)	-0.89	-0.29	-0.22	-0.09	-0.09	0.04	-0.13	0.13	-0.09
BNO (bunch number)	-0.22	-0.60	-0.65	-0.15	-0.08	-0.05	-0.10	0.14	-0.10
ABW (average bunch weight)	-0.96	-0.07	0.13	0.00	-0.05	0.10	-0.10	0.08	-0.01
BWT (bunch weight)	-0.96	-0.10	0.14	-0.10	0.07	0.05	-0.02	0.03	-0.03
MFW (mean fruit weight)	-0.88	0.03	0.35	0.06	-0.04	0.27	-0.04	-0.04	-0.06
MNW (mean nut weight)	-0.21	-0.14	-0.01	0.23	-0.26	0.74	-0.02	0.20	0.04
P/B (parthenocarpic bunch)	-0.83	0.06	0.24	-0.15	0.01	0.11	-0.18	-0.19	-0.15
M/F (mean to fruit ratio)	-0.84	-0.04	0.47	0.03	0.07	0.13	-0.03	-0.15	-0.03
K/F (kernel to fruit ratio)	0.66	-0.06	-0.25	-0.22	-0.14	0.08	-0.32	-0.07	0.49
S/F (shell to fruit ratio)	0.78	0.07	-0.48	0.03	-0.04	-0.18	0.15	0.20	-0.13
O/DM (oil to dry mass ratio)	-0.39	-0.18	0.51	-0.10	-0.01	-0.55	0.14	-0.09	0.09
O/WM (oil to wet mass ratio)	-0.39	-0.20	0.51	-0.09	0.05	-0.53	0.08	-0.21	0.11
F/B (fruit to bunch ratio)	0.24	-0.27	0.44	0.13	-0.34	0.49	0.01	-0.16	0.19
O/B (oil to bunch ratio)	-0.75	-0.16	0.56	0.04	0.05	0.12	0.00	-0.20	0.07
K/B (kernel to bunch ratio)	0.65	-0.16	-0.07	-0.13	-0.25	0.26	-0.28	-0.12	0.52
OY (oil yield)	-0.94	-0.14	0.25	-0.05	0.01	0.11	-0.05	-0.04	-0.02
KY (kernel yield)	-0.30	-0.61	-0.53	-0.31	-0.07	0.14	-0.17	0.06	0.25
TEP (total economic product)	-0.94	-0.24	0.15	-0.10	0.00	0.13	-0.07	-0.03	0.03
FP (frond production)	-0.09	-0.64	-0.27	-0.15	0.00	-0.12	0.39	-0.21	0.22
PCS (petiole cross section)	-0.45	0.78	-0.26	-0.11	-0.11	-0.03	0.07	-0.06	0.12
RL (rachis length)	-0.49	0.74	-0.10	-0.20	0.07	-0.10	-0.11	-0.06	0.05
LL (leaf length)	-0.26	0.43	-0.50	0.17	0.36	-0.08	-0.14	-0.06	0.17
LW (leaf width)	-0.58	0.33	0.12	0.38	-0.38	0.08	-0.08	0.24	-0.13
LN (no of leaves)	-0.47	0.51	0.03	-0.26	0.13	-0.06	0.14	-0.20	0.29
HT (plant height)	0.08	-0.02	-0.26	0.06	-0.31	0.32	0.65	-0.14	-0.22
LA (leaf area)	-0.74	0.60	-0.15	0.17	0.04	0.03	-0.02	-0.01	0.12
LAI (leaf area index)	-0.75	0.59	-0.15	0.17	0.04	0.03	-0.02	-0.01	0.12
DIAM (diameter of trunk)	-0.04	0.72	0.07	0.02	-0.52	-0.11	-0.10	-0.08	0.06
F (fractional rad. intercept.)	-0.63	0.67	-0.16	0.15	-0.02	-0.06	-0.03	0.03	0.17

LAR (leaf area ratio)	-0.39	-0.57	0.11	0.31	0.50	0.07	-0.15	0.07	0.19
BDM (bunch dry mass)	-0.88	-0.32	-0.26	-0.16	0.00	0.02	-0.05	0.09	-0.06
VDM (vegetative dry mass)	-0.38	0.69	-0.35	-0.12	-0.31	-0.03	0.26	-0.17	0.16
TDM (total dry mass)	-0.85	0.18	-0.39	-0.19	-0.18	0.00	0.12	-0.04	0.05
BI (bunch index)	-0.59	-0.68	-0.26	-0.16	0.19	0.03	-0.12	0.15	-0.03
E (radiation efficiency)	-0.69	-0.17	-0.46	-0.40	-0.26	-0.03	0.19	-0.07	-0.04
NAR (net assimilation ratio)	-0.25	-0.50	-0.43	-0.55	-0.31	-0.09	0.20	-0.08	-0.11
C14:0 (myristic acid)	-0.29	-0.49	0.42	0.39	-0.01	-0.17	0.22	0.30	0.06
C16:0 (palmitic acid)	-0.17	-0.30	-0.41	0.76	-0.10	-0.09	0.02	-0.28	0.03
C16:1 (palmitoleic acid)	-0.09	0.17	-0.55	0.45	0.30	-0.09	-0.04	0.02	-0.12
C18:0 (stearic acid)	0.30	0.00	0.41	-0.29	-0.30	-0.03	-0.44	-0.25	-0.35
C18:1 (oleic acid)	0.16	0.43	0.11	-0.73	0.33	0.20	0.05	0.24	-0.04
C18:2 (linoleic acid)	-0.22	-0.44	0.27	0.49	-0.41	-0.28	0.12	0.04	0.25
C18:3 (linolenic acid)	-0.12	0.19	-0.63	0.31	-0.02	-0.06	-0.04	0.22	0.02
C20:0 (arichidic acid)	-0.05	-0.34	-0.29	-0.24	-0.30	-0.33	-0.44	-0.10	-0.04
IV (iodine value)	-0.06	0.16	0.46	-0.50	0.00	-0.06	0.26	0.52	0.29
Carotene	0.06	-0.13	-0.09	-0.12	0.56	0.34	0.19	-0.51	0.01
Eigenvalue	11.20	6.67	5.01	3.47	2.36	2.13	1.65	1.38	1.21
Variance (%)	28	17	13	9	6	5	4	3	3
Cumulative (%)	28	45	58	67	73	78	82	85	88

The scree plot as shown in Figure 1 is a visual aid for determining an appropriate number of PCs. It shows the eigenvalue against the component number. The eigenvalues measure the amount of variation explained by each PC and will be largest for the first PC and smaller for the subsequent PCs. An eigenvalue of greater than one indicates PCs accounted for more variance than accounted by one of the original variables in standardized data and it is commonly used as a benchmark for which PCs are retained [22]. The scree plot in this study showed that maximum variation was present in the first PC and the selection of genotypes from PC 1 will be useful [23].

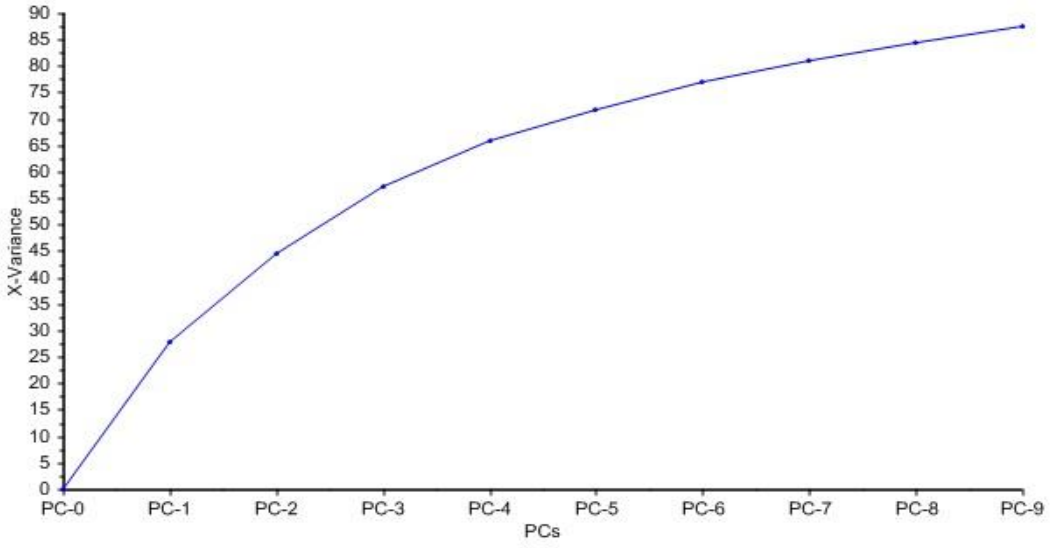

Fig. 1 Scree plot of principal component analysis between percentage variance and number of principal components

B. Loadings

PC loadings (Figures 2 and 3) are the correlation coefficients between the PC scores and the original variables. PC loadings measure the importance of each variable in accounting for the variability in the PC [9]. In this study, variables on the left quadrant with high loadings on PC 1 include TDM, P/B, MFW, ABW, BWT, OY, M/F, OY, TEP, FFB, BDM, O/B and *e* while those on the right quadrant with high loadings include S/F, K/F and K/B; these set of variables can be said to anti-correlated. Those on the left quadrant are traits conferring high yield to palms while those on the right quadrant are not yield contributing traits and oil palms with high percentage of S/F, K/F and K/B will be low in yield. This is evident from the strong negative correlation that was unveiled between

variables and FFB and other yield traits in the research carried out by our co-workers on the association between oil palm traits [24].

Variables with high contribution to PC 2 include BNO, PCS, RL, FP, DIAM and BI. Also, from the loadings plot, variables that lie close together along the same PC can be said to be highly correlated [5]. As stated by previous researcher [9], high correlation between PC 1 and a variable indicates that the variable is associated with the direction of the maximum amount of variation in the data*set also*; more than one variable might have a high correlation between with PC 1. A strong correlation between a variable and PC 2 indicates that the variable is responsible for the next largest variation in the data perpendicular to PC 1 and so on.

Furthermore, on the PC, some variables, particularly amongst the fatty acid traits had no significant loadings on any of the extracted PCs. This suggests that the variables have little or no contribution to the variation in the Senegal and Gambia oil palm germplasm. Therefore, PCA may often indicate which variables in a dataset are important and which ones may be of little importance [9].

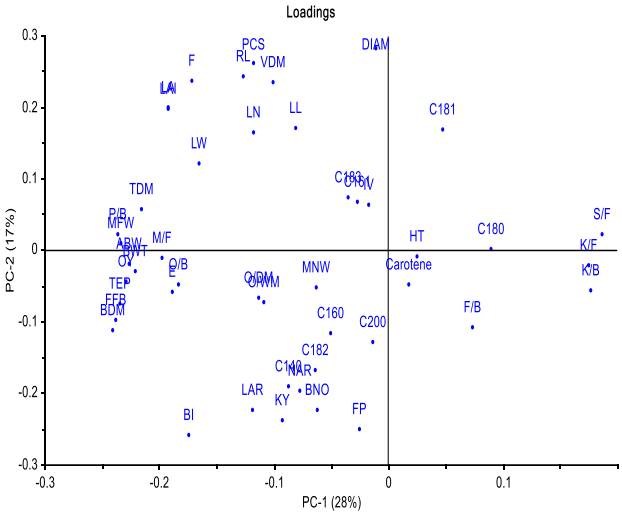

Fig.2 Scattered diagram of 46 oil palm germplasm traits for first two components contributing almost half of the total variability

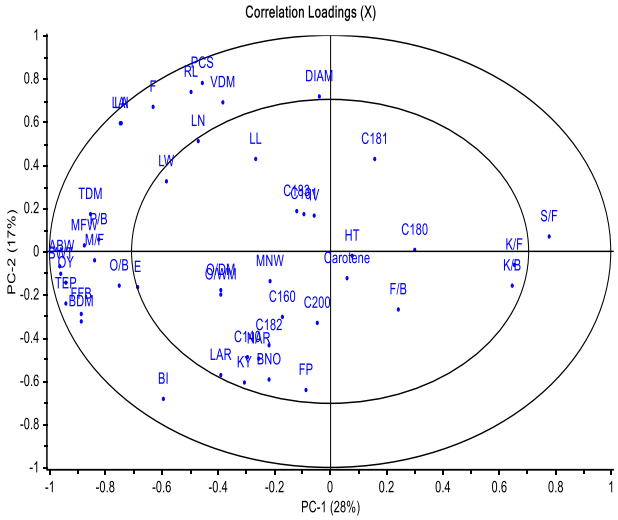

Fig. 3 Scattered diagram of 46 oil palm germplasm traits showing correlation to the first two components

C. Scores

PC scores are the derived composite scores computed for each observation based on the eigenvectors for each PC [9]. From the scores of the oil palm germplasm summarized in Table 2, oil palm genotypes with high scores on PC 1 can be said to be the most diverse. Those with high negative scores i.e. SSC 3, SEN 09.04 and SEN 01.02 can be said to compliment variables with high negative loadings on PC 1 while those with high positive scores i.e. SEN 13.07 and SEN 02.08 are complementary to variables with high positive loadings on PC 1[5]. The oil palm genotypes with high negative scores are of the high yield type as yield traits as observed from the loading plots were loaded negatively on PC 1 while the non-yield traits were also positively loaded on PC 1 and genotypes with high positive scores can be said to be non-yield type. Therefore, from the scores given to the Senegal and Gambia oil palm germplasm, breeders can select genotypes with highest score having desirable characters for further breeding program.

Furthermore, scores plot which is a dimensional scatter plot signifies how well the data is distributed and gives information about the samples. In the scores plot of the present study (Figure 4), oil palm genotypes that are closer to one another have close values of the corresponding variables while those that are far away from one another are quite different in values for corresponding variables [5]. The scores plot of the Senegal and Gambia materials portrayed that genotypes that are close together are sensed as being similar when rated on all the variables studied while genotypes which are further apart are more diverse from other accessions [25].

In the present study, SSC 3 was the most distant oil palm genotype from other oil palm accessions. The reason for this is glaring as SSC 3 is a standard cross; i.e. hybrid between a *dura* and a *pisifera* and standard crops also known as *tenera* are known to be f high commercial value as they are high yielding genotypes as compared to their *dura* counterparts used in this study [26]; [27]. Asides from SSC 3, genotypes which are also diverse as can be seen from the scores plot include SEN 01.02, SEN 09.04 and GAM 05.08. The Gambia

material GAM 05.08 was expected to be different because it is not of the same origin with the “SEN” genotypes.

TABLE II
SCORES OF THE 42 SENEGAL-GAMBIAN OIL PALM GERMPLASM ON THE EXTRACTED PCS

	Principal Components								
Accessions	PC-1	PC-2	PC-3	PC-4	PC-5	PC-6	PC-7	PC-8	PC-9
GAM 05.02	1.29	-0.20	1.23	-1.77	1.03	-0.78	1.67	-2.32	0.97
GAM 05.08	1.79	-5.64	2.19	1.67	2.79	4.07	-1.39	-1.04	1.49
SEN 01.02	-3.11	5.96	-8.83	4.30	1.25	1.33	1.73	1.29	0.23
SEN 01.03	-0.57	0.09	0.49	-0.45	-1.02	0.60	1.53	0.25	-0.44
SEN 02.01	0.12	0.72	-0.75	-0.57	0.45	-0.58	-0.38	-0.29	0.67
SEN 02.04	0.17	0.02	-0.21	-1.44	-1.10	-0.57	0.63	1.56	0.29
SEN 02.05	-0.12	-3.47	-1.48	-1.93	0.10	-0.75	-0.52	-0.28	0.43
SEN 02.06	0.37	-0.63	-2.06	-1.37	-1.41	-0.31	0.22	0.70	0.25
SEN 02.08	4.25	5.04	3.20	-0.54	1.30	1.89	0.73	-0.05	-1.42
SEN 02.09	0.70	-0.73	-1.73	-0.47	0.54	0.40	-0.07	-0.43	-1.10
SEN 03.03	0.43	0.42	-1.36	0.94	0.33	-0.23	-0.60	-0.12	0.29
SEN 03.06	0.44	-0.49	-1.32	3.93	2.82	-1.35	-0.58	-1.80	-0.70
SEN 03.07	2.63	-1.70	0.20	0.93	2.04	1.45	-1.58	-0.51	-0.62
SEN 04.01	1.48	0.73	0.84	0.39	1.12	-1.74	-1.13	0.50	0.76
SEN 04.02	0.94	-1.58	-1.06	1.46	-1.38	-2.47	-0.79	-0.65	-1.03
SEN 04.03	-0.79	-1.66	-1.64	1.17	0.80	-1.44	-1.34	1.54	1.31
SEN 05.01	0.53	-1.27	1.25	0.60	-0.51	-1.27	-2.31	0.78	0.71
SEN 05.02	-0.06	-3.63	-0.02	0.96	-1.71	3.49	0.43	1.13	1.63
SEN 05.03	-0.83	-1.83	-1.79	-0.14	-0.49	-1.28	-0.68	0.65	0.48
SEN 05.04	0.74	-0.33	-0.37	-1.04	2.51	0.72	-0.94	0.37	-0.18
SEN 05.05	1.55	-0.94	-1.79	-0.99	-0.01	2.34	0.37	0.36	-0.38
SEN 05.08	0.61	-2.88	-0.57	-2.47	-0.37	-0.72	0.45	0.59	-2.26
SEN 06.01	0.41	-1.67	-0.71	-1.55	-1.68	-1.54	-0.52	0.23	1.33
SEN 06.08	-0.12	-0.69	0.47	-1.07	0.31	1.80	1.47	0.77	0.08
SEN 07.03	-0.03	-0.68	0.91	-1.62	-0.44	0.01	0.59	0.40	1.08
SEN 07.04	-0.49	-1.73	-1.59	-1.06	-1.38	-0.12	0.51	-1.37	-1.07
SEN 07.05	2.21	-1.77	1.80	3.19	0.00	-1.11	1.15	-1.94	-0.16
SEN 07.08	0.98	-2.15	-1.96	-1.56	-0.24	-0.02	1.59	0.13	-0.80
SEN 08.02	-0.45	-1.92	-2.83	1.13	-1.76	0.57	-1.87	0.19	-1.45
SEN 08.03	2.87	0.33	3.28	0.21	1.68	0.59	1.25	2.74	-1.17
SEN 08.04	2.81	-1.19	2.32	1.94	0.33	-0.82	3.11	0.27	-1.78
SEN 09.04	-7.87	-1.73	3.43	2.45	0.09	-1.23	1.83	2.33	1.08
SEN 10.03	1.21	-0.77	1.51	2.25	-0.69	-1.26	1.01	-0.50	1.26
SEN 10.05	1.13	-2.22	-0.02	-1.48	-1.16	-0.80	0.78	-0.53	-1.23
SEN 12.01	2.07	2.37	0.61	-0.19	-0.63	-0.91	0.03	-0.60	0.59
SEN 12.02	2.42	4.12	-0.27	-3.60	3.02	-1.15	-2.41	1.38	-1.04
SEN 12.03	0.92	3.50	0.38	0.44	1.31	-1.83	1.00	-2.07	1.28
SEN 12.05	2.77	3.90	1.24	-2.21	-0.40	0.30	-0.09	0.22	2.56
SEN 13.01	2.45	2.12	-2.19	-0.95	-3.10	2.52	-0.12	-2.35	0.44
SEN 13.04	1.30	4.57	0.52	-1.31	-0.79	0.21	0.87	-0.14	0.99
SEN 13.07	4.83	4.23	3.91	4.31	-4.40	0.65	-2.49	0.92	-1.09
SSC 3	-16.53	3.01	3.46	-1.12	-0.23	1.23	-1.32	-1.60	-1.22

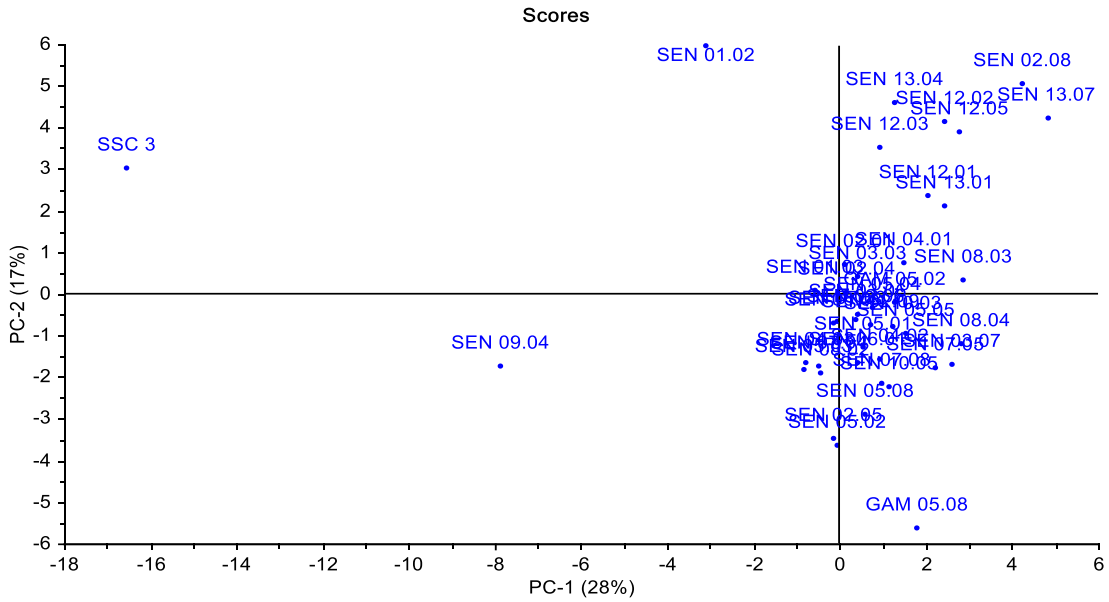

Fig. 4 Two dimensional ordinations of 42 Senegal-Gambian oil palm accessions on principal axes 1 and 2

D. Bi-Plot

The Bi-plot display is a visualization technique for investigating the inter-relationships between the observations and variables in multivariate data [9]. From the bi-plot (Figure 5), genotypes SSC 3, SEN 09.04 and SEN 01.02 will be a good choice for genetic improvement. GAM 05.08 will be the best choice for high carotene palm. The result of this is in conformity with that of Doumbia *et al*. [11] who used the bi-plot graph to suggest good candidates of cowpea accessions to be used in genetic improvement of the crop.

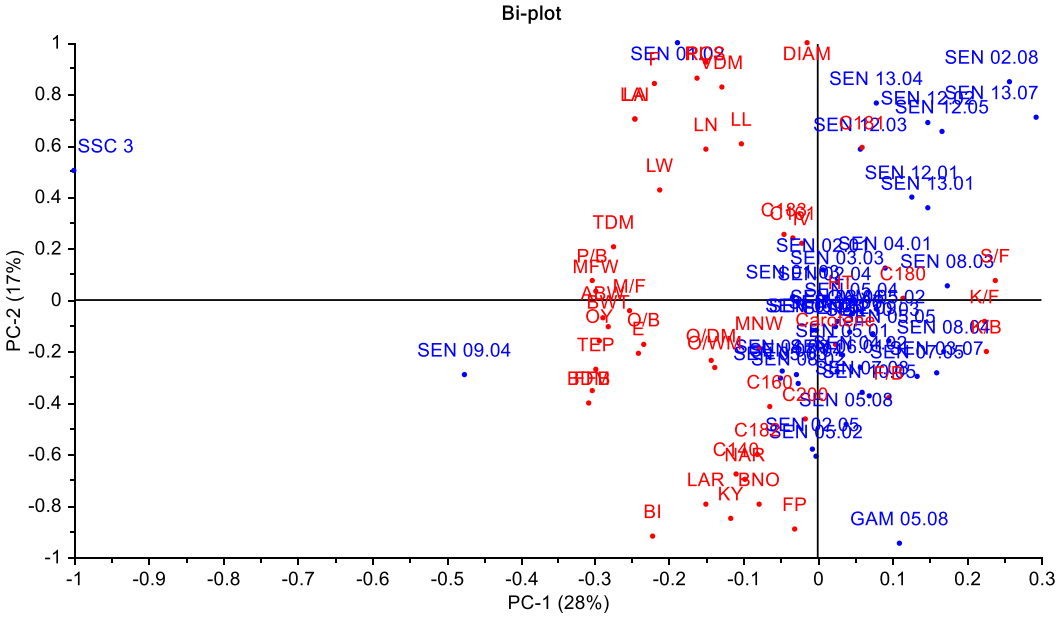

Fig. 5 Bi-plot of 46 oil palm agronomic traits and 42 oil palm accessions on PC 1 and PC 2

E. Cluster Analysis

Estimating genetic diversity and determining the relationships between collections are very useful for ensuring efficient germplasm collections and different markers are available for studying the variability among accessions [23]. Several techniques have also been used to classify and measure the patterns of phenotypic diversity in the relationships of species and germplasm collections for a variety of crops. However, morphological characterization constitutes the first step in the description and classification of germplasm [28].

PCA alone may not give a clear character representation in terms of their contribution to genetic diversity and hence, the need for PCA to be complemented with other techniques such as cluster analysis which provides more information about the relative positions of the accessions [8]. Cluster analysis of the germplasm resources is helpful for parental selection in the plant breeding program [10]. In this study, all the Senegal and Gambia oil palm germplasm were clustered into six types with each group having its own peculiar characters (Figure 6). Such information would be effective in selecting parental materials to breed new expected oil palm varieties. The MPOB-Nigerian oil palm germplasm also grouped into eight types by cluster analysis.

Though cluster analysis was able to group accessions with greater morphological similarity together, the grouping did not necessarily group accessions from the same origin together. This can be observed in the grouping of Gambia materials which were not in the same cluster group but were found grouped together with some of the Senegal materials. This shows that there is no consistency between geographical origin and genetic distance. Previous researchers also reported lack of relationship between geographical origin and distance [8] [12]. It is believed that the association between genetic similarity and geographic distance among genotypes is not always clear [28]. This may be due to migration of the oil palm materials from one region to another in collection site.

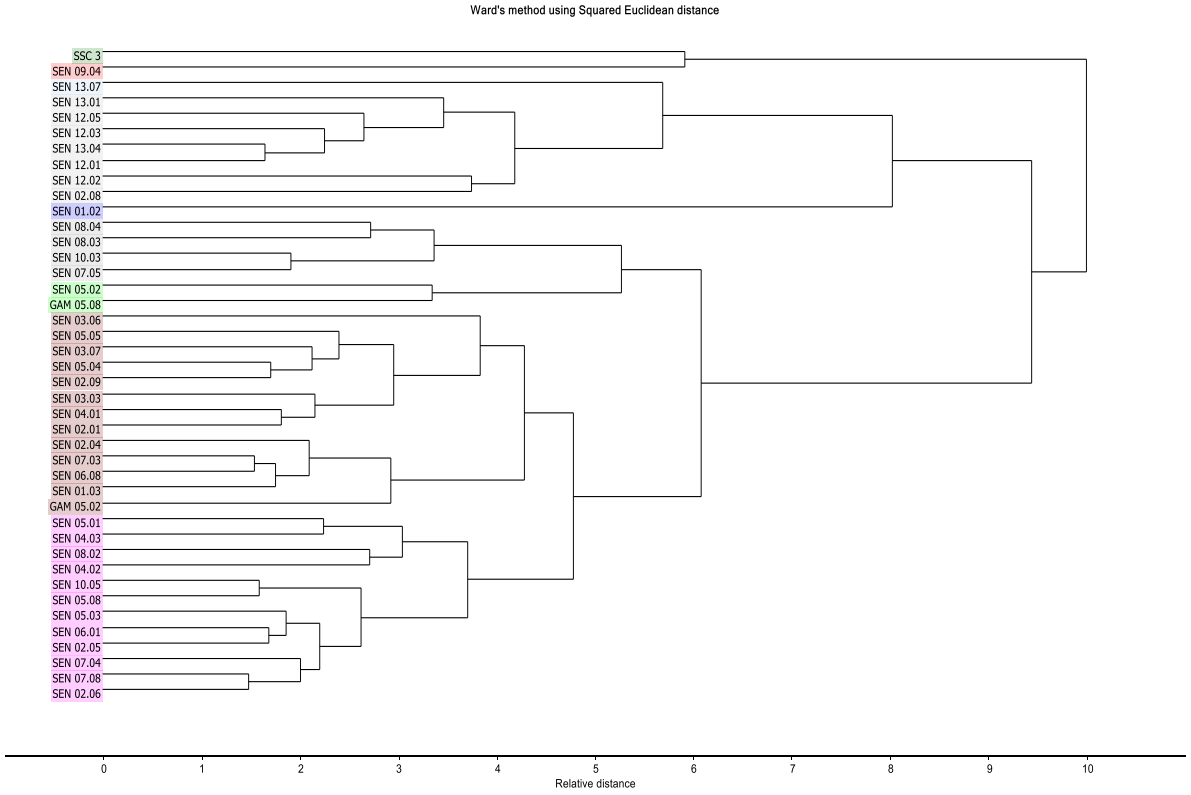

Fig. 6 The relationship among the oil palm germplasm reflected by cluster analysis

Based on the means value for each cluster (Table 3), cluster-I, cluster-II and cluster-III contained genotypes with high yield characters and in line with the goal of plant breeders [8]. Hence, these genotypes could be exploited for their release as high yielding accessions after testing them on a wide range of environments. Furthermore, these genotypes can also be utilized as parents in hybridization programs to develop high yielding oil palm varieties. This finding is in conformity with research of Ajmal *et al*. [8]. Also, cluster-II had the lowest height and hence, the genotype in this group could be used for breeding of short palms as it is a preferred trait in oil palm breeding for easy harvesting of fresh fruit bunch [25]. Groups with desired traits can also be exploited directly for such traits.

TABLE III

CHARACTERISTICS MEANS OF SIX CLUSTERS GENERATED BY WARD'S CLUSTER ANALYSIS BASED ON 46 AGRONOMIC TRAITS

	Cluster-I	Cluster—II	Cluster-III	Cluster-IV	Cluster-V	Cluster-VI
FFB	106.04	141.68	**197.68**	49.15	62.51	88.51
BNO	**21.00**	18.88	15.86	14.22	16.33	20.19
ABW	5.05	8.83	**12.55**	3.37	3.88	4.36
BWT	3.93	8.49	**12.57**	2.63	3.41	3.88
MFW	2.65	6.97	**12.50**	2.31	3.11	2.51
MNW	**1.91**	1.65	1.89	1.53	1.79	1.58
P/B	0.00	0.55	**5.10**	0.10	0.24	0.37
M/F	28.09	57.40	**84.31**	33.80	38.67	35.12
K/F	13.29	9.03	6.41	**16.70**	14.96	15.54
S/F	**58.62**	33.58	9.28	49.50	46.37	49.34
O/DM	61.73	75.15	**77.37**	68.49	70.43	70.33
O/WM	32.13	46.11	**49.68**	39.58	41.66	41.42
F/B	50.76	57.35	55.60	58.20	**60.22**	57.33
O/B	4.58	15.50	**23.12**	7.79	10.21	8.51
K/B	6.78	5.29	3.62	9.63	**9.02**	8.89
OY	4.86	27.17	**45.76**	3.36	7.12	7.57
KY	7.19	5.67	7.16	4.28	6.03	**7.79**
TEP	9.17	30.57	**50.05**	5.93	10.73	12.25
FP	30.00	**33.50**	28.14	28.80	31.76	31.74
PCS	**29.52**	17.88	27.53	20.13	12.87	16.01
RL	4.45	3.94	**5.00**	4.13	3.47	3.85
LL	**107.20**	88.54	88.86	86.60	80.08	86.44
LW	4.57	4.98	**5.16**	4.10	3.91	3.93
LN	136.00	125.50	**158.07**	130.73	121.34	122.35
HT	3.39	2.60	**2.58**	2.77	3.01	2.81
LA	7.59	6.40	**8.28**	5.24	4.36	4.74
LAI	4.49	3.79	**4.90**	3.10	2.58	2.81
DIAM	0.70	0.69	0.74	**0.76**	0.61	0.66
F	0.86	0.78	**0.88**	0.72	0.65	0.68
LAR	12.36	**17.03**	14.60	11.12	14.17	13.06
BDM	8.32	11.11	**15.81**	3.46	4.95	6.96
VDM	**18.43**	12.77	16.15	13.74	10.00	11.74
TDM	26.75	23.89	**31.96**	17.20	14.95	18.70
BI	0.31	0.45	**0.49**	0.19	0.32	0.37
E	1.00	0.96	**1.18**	0.77	0.74	0.89
NAR	11.46	11.76	**12.88**	10.81	11.25	13.15
C140	0.30	**1.13**	0.52	0.36	0.65	0.49
C160	**44.76**	39.74	38.69	37.18	40.27	39.40
C161	**0.53**	0.10	0.13	0.12	0.12	0.16
C180	2.20	4.03	5.32	**5.70**	4.89	5.20
C181	43.45	41.52	45.07	**47.42**	42.85	44.56
C182	8.16	**13.29**	9.94	8.93	10.91	9.86
C183	**0.60**	0.10	0.18	0.17	0.18	0.19
C200	0.00	0.07	0.12	0.10	0.08	**0.14**
IV	53.58	**59.09**	56.58	56.82	56.34	56.05
Carotene	1575.49	1421.42	1607.71	1624.35	**1793.83**	1638.94

Figures in bold are maximum values

F. Genetic Distance

According to squared Euclidean distances (D^2) among the genotypes (Table 4), the largest genetic distance was between SEN 13.07 and SSC 3; this was followed by SEN 02.08 and SSC 3. Hence, crosses between morphologically distant genotypes will result in maximum heterosis. The importance of genetic diversity to maximum heterosis has been reported by many researchers [28].

Genotypes with the largest genetic distance between them can be hybridized as hybrids of maximum distance result in high yield. On the other hand, the least genetic distance was recorded in SEN 02.06 and SEN 07.08. Therefore, crosses between genotypes of proximity should be avoided. However, it was suggested that crosses between close genotypes could be useful for backcross breeding programs [29], [30].

TABLE IV
INTER CLUSTER DISTANCE AS ANALYZED BY PROXIMITY MATRIX OF SQUARED EUCLIDEAN DISTANCE

Case	Squared Euclidean Distance					
	1: Cluster-I	2: Cluster-II	3: Cluster-III	4: Cluster-IV	5: Cluster-V	6: Cluster-VI
1: Cluster-I	0.000	118.586	149.386	99.442	107.001	86.122
2: Cluster-II	118.586	0.000	74.408	91.789	63.147	56.828
3: Cluster-III	149.386	74.408	0.000	151.442	158.627	126.870
4: Cluster-IV	99.442	91.789	151.442	0.000	38.770	36.825
5: Cluster-V	107.001	63.147	158.627	38.770	0.000	20.757
6: Cluster-VI	86.122	56.828	126.870	36.825	20.757	0.000

IV. CONCLUSIONS

Morphological diversity among the Senegal and Gambia oil palm germplasm was well defined by both principal component and clustering analyses. Considering the different morpho-bio-agronomic descriptors, it has been possible to observe a noteworthy inter and intra-group diversity. This recommends the likelihood of attaining, through selection, suitable genotypes combining the high yield with desirable traits for direct release as cultivars in the Malaysian Palm Oil Board (MPOB).

ACKNOWLEDGMENT

This research was supported by Universiti Sains Islam Malaysia (USIM). The authors also express their profound gratitude and thanks to the Malaysian Palm Oil Board (MPOB) for providing the data for this study.

REFERENCES

[1] R.H.V. Corley and P.B. Tinker, *The Oil Palm,* 4th Edition, Wiley-Blackwell, UK, 2003.

[2] N. Rajanaidu, *Oil palm cultivation and FFB production in Selected Readings on Palm Oil and Its Uses,* Palm Oil Research Institute of Malaysia, Kuala Lumpur, Malaysia, 1994.

[3] B.S. Jalani, "Research and development of oil palm towards the next millennium," 1998 International Oil Palm Conference. IOPRI and GAPKI, Bali, Indonesia, 1998.

[4] Oil world, *Oil world 2007 annual report*, Mielke, Hamburg, 2008.

[5] N. Rajanaidu, B.S. Jalani, A. Kushairi, V. Rao, *Oil palm genetic resources-collection, evaluation, utilization and conservation*, In Rajanaidu, N., and B.S. Jalani (Editors) Proceeding of the symposium on the science of oil palm breeding, PORIM, Bangi, Malaysia, 1999.

[6] P.K. Mandal. *Bunch Analysis of Oil Palm*. National Research Centre for Oil Palm (NRCOP), Padavegi, 2008.

[7] J. Martinez-Calco, A.D. Gisbert, M.C. Alamar, R. Hernandorena, C. Romero, G. Ilacer, M.L. Badenes, " Study of a germplasm collection of Loquat (*Eriobotrya japonica* Lindi) by multivariate alaysis," *Genetic Resources and Crop Evolution,* vol. 55, no. 5, pp. 697-703, 2005.

[8] S. Ajmal, N.M Minhas, A. Hamdani, A. Shakir, M. Zubair, M. Ahmad,"Multivariate Analysis of Genetic Divergence in Wheat (*Triticum aestivum*) Germplasm," *Pakistan Journal of Botany,* vol. 45, no. 5, pp. 1643 – 1648, 2013.

[9] M.K. Bozokalfa, D. Esiyok, K. Turhan, "Patterns of Phenotypic Variation in a Germplasm Collection of Pepper (*Capsicum annum* L.) from Turkey," *Spanish Journal of Agricultural Research,* vol. 7, no. 1, pp. 83 – 95, 2009.

[10] Z. Deyong, "Analysis among Main Agronomic Traits of Spring Wheat (*Triticum aestivum*) in Qinghai Tibet Plateau," *Bulgarian Journal of Agricultural Science,* vol. 17, no. 5, pp. 615 – 622, 2011.

[11] I.Z. Doumbia, R. Akromali, J.Y. Asibuo, "Comparative Study of Cowpea Germplasms Diversity from Ghana and Mali using Morphological Characteristics," *Journal of Plant Breeding and Genetics,* vol. 1, no. 3, pp. 139 – 147, 2013.

[12] G. Lacis, V. Trajkovski, I. Rashid, "Phenotypical Variability and Genetic Diversity within Accessions of the Swedish Sour Cherry (*Prunus cerasus* L.) Genetic Resources Collection," *Biologija,* vol. 56, no (1-4), pp. 1-8, 2010.

[13] M. Lohani, D. Singh, J.P. Singh, "Genetic Diversity Assessment Through Principal Component Analysis in Potato (*Solanum tuberosum* L.)," *Vegetable Science,* vol. 39, no.2, pp. 207-209, 2012.

[14] A. Muhammad, S.A. Jatoi, T. Rafique, and G. Abdul, "Genetic Divergence in Indigenous Spinach Genetic Resources for Agronomic Performance and Implication of Multivariate Analyses for Future Selection," *Science, Technology and Development,* vol. 32, no. 2, pp. 7-15, 2013.

[15] H. Martens and Naes, T., *Multivariate Calibration*, Wiley, New York.

[16] G. Blaak, "Oil Palm Prospection Four in the Bamenda Highlands of West Cameroon," Internal report, Unilever, London, 1967.

[17] G.R. Squire, "A Physiological Analysis of Oil Palm Trials," *PORIM Buletin,* vol. 12, pp. 12–31, 1986.

[18] R.E. Timms, "Artefact Peaks in the Preparation and Gas Chromatographic Determination of Methyl Esters," *Australian Journal of Dairy Technology*, pp. 4-5, 1978.

[19] CAMO. (2014) The Unscrambler Methods. [Online]. Available: http://www.camo.com/resources/methods.html

[20] J.H. Ward, "Hierarchical Grouping to Optimize an Objective Function," *Journal of American Statistical Association,* vol. 58, pp. 236-244, 1963.

[21] M. Maqbool, A. Asgar, A.G. Peter, "Effect of Cinnamon Oil On Incidence Of Anthracnose Disease And Postharvest Quality Of Bananas During Storage," J. Agric. Biol., vol. 12, no. 4, 2010.

[22] G. Fernandez. (2010) Principal component analysis. [Online]. Available: http://www.cabnr.unr.edu/saito/classes/ers701/pca2.pdf

[23] M.A. Rabbani, A. Iwabuchi, Y. Murakami, T. Suzuki, K. Takayanagi, " Genetic Diversity of Mustard (*Brassica juncea* L.) Germplasm from Pakistan as Determined by RAPDs," *Euphytica,* vol. 103, no. 2, pp. 235-242, 1998.

[24] M.A. Li-Hammed, A. Kushairi, N. Rajanaidu, H. Mohd Sukri, C.W.N. Che Wan Zanariah, B.S. Jalani, "Variation and Correlation in the MPOB-Nigerian Oil Palm (*Elaeis guineensis* Jacq.) Germplasm," *Proceedings of the 13th Symposium of the Malaysian Society of Applied Biology*, Cherating, Pahang, Malaysia, 8-10 June, 2014.

[25] F.A. Smiullah, A. Khan, U. Ijaz, R. Iftikher, "Genetic diversity assessment in sugarcane using principal component analysis (PCA)", *International Journal of Modern Agriculture,* vol. 2, no. 1, pp. 34-38, 2013.

[26] C.H. Teoh, *The palm oil industry in Malaysia: from seed to frying pan*. Wild World Fund (WWF), Malaysia, 2002.

[27] M. Khodadadi, M.H. Fotokian, M. Miransari, "Genetic Diversity of wheat (*Triticum aestivum* L.) Genotypes Based on Cluster and Principal Component Analyses for Breeding Strategies," *Australian Journal of Crop Science,* vol. 5, no. 1, pp. 17-24, 2011.

[28] G. WeldeMicheal, S. Alamerew, T. Kufa, T. Benti, "Genetic Diversity Analysis of Some Ethopian Specialty Coffee (Coffea Arabica L.) Germplasm Accessions Based on Morphological Traits," *Time Journals of Agriculture and Veterinary Sciences,* vol. 1, no. 4, pp. 47-54, 2013.

[29] G. Sonnante, D. Pignone, "The Major Italian landraces of Lentil (*Lens culinaris* Medik): Their Molecular Diversity and Possible Origin," *Genetic Resources of Crop Evolution,* vol. 54, pp. 1023-1031, 2007.

[30] M.A. Rahim, A.A. Mia, F. Mahmud, N. Zeba, K. Afrin, "Genetic Variability, Character Association and Genetic Divergence in Mungbean (*Vigna radiate* L. Wilczek)," *Plant Omic,* vol. 3, pp. 1-6, 2010.