CHAPTER III

MATERIALS AND METHODS

3.1 Breeding Materials and Site Location

The germplasm used in this study originated from Senegal and Gambia. A random sample of five families from 13 sites originating from Senegal and also Gambia (population; designated as SEN 01, SEN 02, SEN 03, SEN 04, SEN 05, SEN 05, SEN 06, SEN 07, SEN 08, SEN 09, SEN 10, SEN 11, SEN 12, SEN 13 GAM05.02 and GAM05.08) were collected in July-August 1993 by researchers from Malaysian Palm Oil Board (MPOB) from the two countries. The germplasm materials were planted at 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.

3.2 Data Collection

The performance of progenies for yield and oil yield components was assessed based on data on fresh fruit bunch (FFB) yield records, bunch number (BNO) and average bunch weight (ABW), bunch components, vegetative and physiological characters and fatty acid characters.

UNIVERSITI SAINS ISLAM MALAYSIA
ISLAMIC SCIENCE UNIVERSITY OF MALAYSIA

3.2.1 Yield and Yield Components

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. Data on yield and yield components were evaluated from year 2000 – 2007.

Procedure for Yield Recording

Recordings of bunch weight (BWT) and bunch number (BNO) on individual are carried out during the harvesting rounds in oil palm breeding. Fresh bunch weight (FFB) is the sum of the bunch weight (BWT) while BNO is the total of all the bunch counts and average bunch weight (ABWT) is the quotient between FFB and BNO. The yield components were derived as follows:

$$\text{FFB (kg/p/yr)} = \sum_{i=1}^{n} BWT_i$$

$$\text{BNO (bunches/p/yr)} = \sum_{i=1}^{n} BNO_i$$

$$\text{ABWT (kg/p/yr)} = \text{FFB/BNO}$$

Where n is the number of harvesting rounds.

3.2.2 Bunch Analysis

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. (1963).

Procedure for Bunch Analysis

Samples of three to five bunches from each palm were analysed. Each bunch was weighed and fruit bearing spikelets chopped off the stalk. The spikelets were randomly segmented for analysis of the fruit to bunch (F/B) and fruit compositions (FC). To ease picking, the F/B portions were kept for two days and the weight of empty spikelets, fertile and parthenocarpic fruits were recorded. Analysis of the fruit compositions was continued following spikelet sampling on the same day. Fruits were detached and randomly sampled, counted, weighed and scraped. The nuts were crack opened and kernel weighed. A 5g sample of minced mesocarp was oil extracted in hexane for 18-24 hours using soxhlet apparatus. The amount of oil in the sample was calculated on weight difference before and after extraction. The components of the bunch were calculated using the below formulae:

F/B = Fruit to Bunch (%) = [(FFWT + PFWT) / SWT] × [(BWT -STKWT) / BWT] × 100

P/F = Parthenocarpic to Fruit (%) = PFWT / (FFWT + PFWT) × 100

M/F = Mesocarp to Fruit (%) = [(FSWT - FNWT) / FSWT] × 100

MC = Moisture Content of mesocarp(%) =100 -[{(TDMWT-TWT)/(FSWT-FNWT)}×100]

O/DM = Oil to Dry Mesocarp (%) = [(ETMWT - ETFWT) / (ETMWT - ETWT)] × 100

O/WM = Oil to Wet Mesocarp (%) = [(100 - MC) × O/DM] / 100

O/B = Oil to Bunch (%) = (F/B × M/F × O/WM) / 10,000

O/Fi = Oil to Fibre (%) = [(ETMWT -ETFWT) / (ETFWT -ETWT)] × 100

K/F = Kernel to Fruit (%) = (KWT / FSWT) × 100

MNW = Mean Nut Weight (g) = FNWT /NOFNUT

MFW =Mean Fruit Weight (g) = FSWT / NOFNUT

P/B = Parthenocarpic to Bunch (%) = (P/F × F/B) / 100

OPY = Oil per Palm per Year (kg) = (O/B × FFB) / 100

UNIVERSITI SAINS ISLAM MALAYSIA
ISLAMIC SCIENCE UNIVERSITY OF MALAYSIA

KPY = Kernel per Palm per Year (kg) = (K/B × FFB) / 100

Where;

BWT = Bunch weight

SWT = Spikelet weight

PFWT = Parthenocarpic fruit weight

FSWT = Fruit sub sample weight

NOFNUT = Number of fresh nut

TDMWT = Tin + dry mesocarp weight

ETWT = Extraction thimble weight

ETMWT = Extraction thimble + mesocarp weight

STKWT = Stalk weight

FFWT = Fertile fruit weight

ESPKWT = Empty spikelet weight

FNWT = Fresh nut weight

KWT = Kernel weight

TWT = Tin weight

ETFWT = Extraction thimble + fibre weight

FFB = Fresh fruit bunch

3.2.3 Vegetative Measurements and Physiological Characters

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 (1986). Data on the physiological characters were assessed in the year 2007. The various components of the vegetative and physiological parameters were calculated using the below formulae:

FP = Frond production (no/p/yr) = number of fronds produced in one year

PCS = Petiole Cross Sectional Area (cm^2) = petiole depth × petiole width

RL = Rachis length (m) = length from tip of rachis to the first ligule

LL = Leaflet Length = (LL1 + LL2 + ...LL6) / 6

LW = Leaflet Width = (LW1 + LW2 + ...LW6) / 6

LN = Leaflet number (no/p/yr) = (number of fronds on one side of rachis)×2

D = Trunk diameter (cm) = diameter of trunk at one meter from ground

HT = Trunk height (m) = height of trunk of ground to base of frond 41

HTI = Height increment = (HT at year t) / (age at year t - 2)

LA = Leaf Area (m^2) = $[(\Sigma (LL_i \times LW_i) \times LN \times 0.57) / 6] / 10{,}000$

LAR = Leaf Area Ratio = (FP × LA) / VDM

LAI = Leaf Index Ratio = (40 fronds × LA × PD^{+}) / 10,000

LDW = Leaf Dry Weight (kg) = (0.1023 × PCS) + 0.2062

TDW = Trunk Dry weight (kg) = $3.142 \times (D/2)^2 \times (HT / \text{Age in year}) \times 1{,}000 \times 0.17$

FDW = Frond Dry weight (kg) = FP × LDW

FI = Frond Index = LA / LDW

f = Fractional interception = $1 - \exp^{(-0.47) \times (LAI - 0.3)}$

e = Conversion Efficiency (g/MJ) = TDM / (31 × f)

VDM = Vegetative Dry Matter (kg/p/yr) = (FDW + TDW)

BDM = Bunch Dry Matter (kg/p/yr) = 0.53 × FFB

TDM = Total Dry Matter = [(VDM + BDM) × PD^{+}] / 1,000

BI = Bunch Index = BDM / (BDM + VDM)

NAR = Net Assimilation Rate = TDM / (0.52 × LAI)

TOIL = Total Oil (kg/p/yr) = OPY + (KPY x 0.5*)

TEP = Total Economic Product (kg/p/yr) = OPY + (KPY × $0.6^{\#}$)

Where;

$^{+}$PD = Planting Density
*0.5 = Kernel Oil Extraction Rate
$^{\#}$0.6 = Relative Price of Palm Kernel Oil to Price of Palm Oil
FFB = Fresh Fruit Bunch (kg/p/yr)
OPY = Oil per Palm Year (kg)
KPY = Kernel per Palm per Year (kg)

3.2.4 Fatty Acid Traits

Fatty acid traits were pooled between years 2000 – 2007. The fatty acid composition was evaluated using the method proposed by Timms (1978) for routine analysis using gas chromatography.

Procedure for Fatty Acid Analysis

The fatty acid analysis was carried out using MPOB method Timms (1978). Fresh bunches were collected from the palm, weighed and chopped. Spikelets samples were selected one each from apical, middle and basal portion of the bunch. The samples

UNIVERSITI SAINS ISLAM MALAYSIA
ISLAMIC SCIENCE UNIVERSITY OF MALAYSIA

were sterilized and fruits from the sterilized spikelets separated. The mesocarps were separated and oven dried at 105 °C for a minimum of 3 hours. The dried mesocarp were minced and dried. After mincing, the dried mesocarp was diluted with 300ml hexane and filtered with a spoonful of sodium sulphate. The filtrate was sealed and kept in dry place. Oil and hexane mixture was later separated into oil and hexane, using rotary evaporator by distilling the hexane vapors under vacuum. The extracted oil is then poured in vials, labeled and stored for FAC and carotene analysis.

For the fatty acid methyl determination, 0.05g of the extracted oil was measured and 1.9ml of solvent hexane added and homogenized. Sodium methoxide of 0.1ml was also added until a cloudy solution is formed. After 10 minutes, the clear portion of the methyl ester was pipetted into gas chromatograph vials and covered with Teflon vial caps for analysis. The separation of the FAME by the GC machine was under programmed conditions. The fatty acids components were represented by the peaks on the chromatograph.

3.3 Statistical Analyses

3.3.1 Variability Profile

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 traits were calculated using SPSS statistical tool. This was done in order to know the extent of variation in the germplasm accessions. The average of the quantitative data was also standardized to give equal

UNIVERSITI SAINS ISLAM MALAYSIA
ISLAMIC SCIENCE UNIVERSITY OF MALAYSIA

weight to all measurements using the following formula (Microsoft Office Excel, 2007):

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

Where, X = Value to standardize

μ = Arithmetic mean

σ = standard deviation of the distribution

3.3.2 Principal Components Analysis (PCA)

PCA simplifies the complex data by transforming 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 (CAMO software version 10.1). Mathematically, PCA involved in the decomposition of 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 data is the combination of scores and loadings in which focused by user in interpretation of PCA results while the remaining part is called error or residual matrix. The ath column of T and ath row of P is represented by vectors of t_a and p_a respectively and both are the vector representations of the ath 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 (CAMO, 2014).

In contrast, the residual matrix, *E*, represent the fraction of variation that cannot be modeled well. It other words, it cannot be explained by available PCs, yet, useful in the lack-of-fit measurement of model to the original data where small value of *E* indicated the good model and vice-versa (CAMO, 2014). Usually after PCA, the size of each component can be measured and represented by eigenvalue. In PCA, the more significant the components indicated the larger of their size, thus have larger eigenvalue. Therefore, the eigenvalue of a PC is the sum of squares of the scores and represented as follows:

$$g_a = \sum_{i=1}^{I} tij^2 \tag{3}$$

where g_a is the ath eigenvalue. The sum of all nonzero eigenvalues for a data matrix equals the sum of squares of the entire data-matrix, so that

$$\sum_{a=1}^{k} g_a = \sum_{i=1}^{I} \sum_{j=1}^{J} x_{ij}^2 \tag{4}$$

where *K* is the smaller of *I* or *J* (Brereton, 2003).

3.3.3 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 (CAMO software version 10.1). The general purpose of cluster analysis is to group similar objects or samples into respective classes based on their specified

characteristics or variables and it includes different type of algorithms such as joining (tree clustering), two-way joining (block clustering) and *k*-means clustering. The aim of the tree clustering algorithm was to join the objects or samples in each class of themselves together using specified distance measures and linkage rules, thus, forming larger cluster by connecting all the objects at the last step known as hierarchical tree.

The Ward's method (Ward, 1963) 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:

$$SSE = \sum_{i=1}^{k}\sum_{j=1}^{n_i}(y_{ij} - \bar{y})^2 \qquad (5)$$

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.

UNIVERSITI SAINS ISLAM MALAYSIA
ISLAMIC SCIENCE UNIVERSITY OF MALAYSIA

CHEMOMETRIC ANALYSIS OF OIL PALM (*Elaeis guineensis* JACQ.) GERMPLASM FROM SENEGAL AND GAMBIA

HANA SALEH ALHAJ ALI

UNIVERSITI SAINS ISLAM MALAYSIA
جامعة العلوم الإسلامية ماليزية
ISLAMIC SCIENCE UNIVERSITY OF MALAYSIA
UNIVERSITI SAINS ISLAM MALAYSIA.

CHEMOMETRIC ANALYSIS OF OIL PALM (*Elaeis guineensis* Jacq.) GERMPLASM FROM SENEGAL AND GAMBIA

Hana Saleh Alhadi Abdarhman

(Matric No. 3130043)

Dissertation submitted in partial fulfillment for the degree of

MASTER OF SCIENCE

(FOOD BIOTECHNOLOGY)

Faculty of Science and Technology

UNIVERSITI SAINS ISLAM MALAYSIA

Nilai

May 2015

UNIVERSITI SAINS ISLAM MALAYSIA
ISLAMIC SCIENCE UNIVERSITY OF MALAYSIA

DECLARATION OF THESIS AND COPYRIGHT

إقرار البحث العلمي وحق النشر

Student's Full Name / اسم الطالب بالكامل	Hana Saleh Alhadi		
Academic Session / الفصل الدراسي	2015	Matric No. / رقم التسجيل الجامعي	313643
Research Title / عنوان البحث	Chemometric analysis of oil Palm (Elaeis guineensis Jacq.) germplasm from Senegal and gambia.		

I hereby declare that the work in this thesis/ project paper is my own except for quotations and summaries which have been duly acknowledged /

إني أقر وأعترف بأن هذا البحث من عملي وجهدي الشخصي، أما المقتطفات والاقتباسات، فقد أشرت إلى مصادرها في هامش البحث

I acknowledged that Universiti Sains Islam Malaysia reserves the right as follows /

أقر بأن تحتفظ جامعة العلوم الإسلامية الماليزية بالحقوق التالية:

1. The thesis/ postgraduate project paper is the property of Universiti Sains Islam Malaysia / إن ملكية هذا البحث العلمي لجامعة العلوم الإسلامية الماليزية

2. The library of Universiti Sains Islam Malaysia has the right to publish my thesis/ postgraduate project paper as online open access (fulltext) and make copies for the purpose of research or teaching and learning only /

تحتفظ مكتبة جامعة العلوم الإسلامية الماليزية بحق النشر لهذا البحث العلمي أو المشروع العلمي على الإنترنت (البحث الكامل) وحق الطباعة بهدف إجراء البحث العلمي أو التعليم والتعلم فقط

Hana
(Student's Signature)

245479
(IC No./Passport No.)

1-9-2015
(Date)

(Supervisor's Signature)

DR Mohd Sukri Hassan
(Name)

1.9.2015
(Date)

DR. MOHD SUKRI BIN HASSAN
Director
Institute of Halal Research & Management (IHRAM)
Universiti Sains Islam Malaysia

UNIVERSITI SAINS ISLAM MALAYSIA
ISLAMIC SCIENCE UNIVERSITY OF MALAYSIA
جامعة العلوم الإسلامية الماليزية

AUTHOR DECLARATION

I hereby declare that the work in this dissertation is my own except for quotations and summaries which have been duly acknowledged.

Date: 15th May, 2015

Signature:

Name: Hana Saleh Alhadi Abdarhman

Matric No: 3130043

Address: P6-11A-05 Plaza Indh

Sepakat Indah1, Kajang

43000, Selangor

UNIVERSITI SAINS ISLAM MALAYSIA
ISLAMIC SCIENCE UNIVERSITY OF MALAYSIA

BIODATA OF AUTHOR

Hana Saleh Alhadi (Matric No. 3130043) was born on 8th February, 1988. She is a Libyan by Nationality. She is currently living at P6-11A-05 Plasa Indah Apartment Sepkat Indah 1, Kajang, Selangor, Malaysia. She obtained her Bachelor of Science (Botany) from 7th October University, Bani Walid, Libya. Being an enthusiastic person, with the purpose of exploring knowledge and skill of the field she enrolled in September, 2013 for her Master Degree in Food Biotechnology at Universiti Sains Islam Malaysia, Bandar Baru Nilai, Malaysia. She can be contacted by email via (hs.zbida@yahoo.com).

USIM
UNIVERSITI SAINS ISLAM MALAYSIA
ISLAMIC SCIENCE UNIVERSITY OF MALAYSIA

ACKNOWLEDGEMENTS

My first and foremost thanks is to Allah (SWT) who has made all things possible. I would like to express my gratitude to my supervisor Emeritus Professor Dr. Jalani Sukaimi for the useful comments, remarks and engagement through the learning process of this master dissertation. Your advice on both research as well as on my career have been priceless. Furthermore, I would like to thank my co-supervisor, Dr. Mohd Sukri Hassan for his supervisor on chemometrics for my thesis.

Also, I would like to thank staff of the Faculty of Science and Technology of the university; the Dean of the faculty, Prof. Bachok Taib, staff of Food Biotechnology, laboratory assistants and technicians of the faculty. I am thankful for their aspiring guidance, invaluably constructive criticism and friendly advice during my study. I am sincerely grateful to them for sharing their truthful and illuminating views on a number of issues related to the research and career.

A special thanks to my family. Words cannot express how grateful I am to my mother, father, brothers and sisters for all of the sacrifices that you've made on my behalf. Your prayer for me was what sustained me thus far. At the end I would like express appreciation to the Government of Libya for financial support of my studies. Without you, my appreciation would not be completed.

USIM
UNIVERSITI SAINS ISLAM MALAYSIA
ISLAMIC SCIENCE UNIVERSITY OF MALAYSIA

ABSTRAK

Anggaran kepelbagaian genetik dan menentukan hubungan antara koleksi adalah strategi yang penting untuk memastikan pengumpulan dan penggunaan germplasma yang cekap. Bahan germplasma sawit yang dikutip dari Senegal dan Gambia yang ditanam di Stesen Kluang MPOB telah dicirikan mengikut kepelbagaian genetik. Sebanyak 44 ciri - ciri agronomi germplasma sawit ini telah dinilai melalui statistik yang mudah untuk menilai kepelbagaian genetik; dan dua teknik kemometrik (PCA dan analisis Kluster) untuk mengenal pasti sifat - sifat yang menyumbang kepada perubahan keseluruhan dan mengklasifikasikan bahan-bahan itu berdasarkan persamaan. Keputusan profil kepelbagaian itu menunjukkan bahawa germplasma kelapa sawit Senegal dan Gambia mempunyai kebole hubahan yang rendah kepada tinggi untuk pelbagai ciri. Sembilan komponen utama dengan nilai eigen > 1 bersamaan dengan 88% daripada jumlah variasi dengan majoriti variasi itu dikuasai oleh PC1. Kebanyakan ciri-ciri, khususnya, BTS, ABW, BWT, MFW, P / B, M / F, OY, TEP, TDM, BDM, e, S / F, K / F, O / B dan K / B menyumbang kepada kepelbagaian di antara dan dalam germplasma, yang mana menunjukkan bahawa variasi yang luas wujud dalam bahan germplasma yang telah dikaji. Analisis kumpulan Ward adalah berdasarkan keputusan - keputusan PCA yang telah dikumpulkan untuk aksesi 42 minyak kelapa sawit kepada enam kelompok, dan kelompok-VI mempunyai bilangan tertinggi ahli. Tambahan pula, ia tiada kaitan antara kepelbagaian genetik dan asal geografi. Purata ciri-ciri agronomi setiap kelompok menunjukkan bahawa kelompok-III mempunyai nilai purata tertinggi pada sifat hasilnya. Selain itu, kumpulan - kumpulan kelompok yang mempunyai nilai purata yang tinggi bagi ciri-ciri yang diingini boleh dipilih untuk ciri-ciri setiap satu. Pengaksesan ini boleh digunakan untuk menghasilkan bahan-bahan kelapa sawit yang berhasil tinggi.

Kata kunci: Kelapa sawit, germplasma, kepelbagaian genetik, kimometrik, analisis komponen utama, analisis kelompok.

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 MPOB Kluang Station were characterized for genetic diversity. A total of 44 agronomic traits of these oil palm materials were subjected to simple statistics to evaluate the genetic variability; and two chemometric techniques (PCA and Cluster analysis) to identify the characters contributing to the overall variation and classify the materials based on similarity. 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 PC1 capturing majority of the variation. Most of the traits especially, FFB, ABW, BWT, MFW, P/B, M/F, OY, TEP, TDM, BDM, *e*, S/F, K/F, O/B and K/B contributed to 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. 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, germplasm, genetic diversity, chemometrics, principal component analysis, cluster analysis.

UNIVERSITI SAINS ISLAM MALAYSIA
ISLAMIC SCIENCE UNIVERSITY OF MALAYSIA
YSIA

ملخص البحث

تقدير التنوع الوراثي وتحديد العلاقات بين المجموعات هي استراتيجيات مفيدة لضمان جمع المادة الوراثية والاستفادة منها. تميزت مواد الأصول الوراثية لنخيل الزيت التي تم جمعها من السنغال وغامبيا من محطة MPOB Kluang في الحفاظ علي التنوع الجيني. وتعرض 44 من الصفات الزراعية لزيت النخيل لإحصاءات بسيطة لتقييم التباين الوراثي.

وتقنيات تحليل المكون الرئيسي (PCA) والتحليل العنقودي (CA) للتعرف على الشخصيات المساهمة في الاختلاف الكلي وتصنيف المواد على أساس التشابه. وأظهرت نتائج تحليل المكون الرئيسي (PCA) أن الأصول الوراثية لنخيل الزيت من السنغال وغامبيا عرضت تفاوت كبير لمختلف الصفات. تسعة عناصر رئيسية مع القيمة الذاتية > 1 تمثل 88 من التباين الكلي وأغلب الاختلاف من PC1. ساهم معظم الصفات خاصة، MFW, BWT, ABW, FFB, BDM, TDM, FE,P OY, M /, F P / B, O / B, K / F, S / F, وK / B إلى الاختلاف بين المادة الوراثية، مشيرا إلى أن تباين واسع موجود في مواد الأصول الوراثية المدروسة. التحليل العنقودي (CA) وارد بناء على نتائج تحليل المكون الرئيسي (PCA) حيث تم تقسيم 42 من نخيل الزيت الي ست مجموعات. كتلة 5 تميزت بوجود أكبر عدد من الأعضاء. وعلاوة على ذلك لم يكن هناك ارتباط بين التنوع الجيني والأصل الجغرافي. وأظهرت الصفات الزراعية أن الكتلة 3 تميزت بوجود أعلى قيمة لمتوسط العائد من الصفات أيضا. ويمكن استخدام هذه المجموعات لإنتاج زيت النخيل ذات بعوائد عالية.

الكلمات المفتاحية للبحث : الأصول الوراثية لزيت النخيل، التنوع الجيني، chemometrics، تحليل المكون الرئيسي و التحليل العنقودي.

UNIVERSITI SAINS ISLAM MALAYSIA
جامعة العلوم الإسلامية الماليزية
ISLAMIC SCIENCE UNIVERSITY OF MALAYSIA

CONTENT PAGE

Contents **Page**

AUTHOR DECLARATION	i
BIODATA OF AUTHOR	ii
ACKNOWLEDGEMENTS	iii
ABSTRACT	v
ABSTRACT	vi
ABSTRACT	vii
CONTENT PAGE	viii
LIST OF TABLES	xi
LIST OF FIGURES	xii
LIST OF APPENDICES	xiii
ABBREVIATION	xiv
CHAPTER I: INTRODUCTION	1
1.1 Problem Statement	2
1.2 Aim	3
1.3 Research Questions	3
1.4 Objectives	3
CHAPTER II: LITERATURE REVIEW	4
2.1 HISTORY OF THE GENUS ELAEIS	4
2.1.1 Evolution of the Genus *Elaeis*	5
2.1.2 Morphology and Biology of *E. guineensis*	6
2.1.3 Infra Specific Classification of *E. guineensis*	7
2.1.4 Ecology and Morphology of *E. guineensis*	9
2.2 PLANT GENETIC RESOURCES	13

UNIVERSITI SAINS ISLAM MALAYSIA
جامعة العلوم الإسلامية الماليزية
ISLAMIC SCIENCE UNIVERSITY OF MALAYSIA

2.2.1 Genetic Diversity	13
2.2.2 Characterization of Genetic Diversity	14
2.3 CHEMOMETRICS	15
2.3.1 Principal Components Analysis (PCA)	16
2.3.1 Clustering Analysis	18
2.3.1.1 Types of Clustering	20
2.4 PREVIOUS STUDIES	21
CHAPTER III: MATERIALS AND METHODS	27
3.1 Breeding Materials and Site Location	27
3.2 Data Collection	27
3.2.1Yield and Yield Components	27
3.2.2 Bunch Analysis	28
3.2.3 Vegetative Measurements and Physiological Characters	30
3.2.4 Fatty Acid Traits	31
3.3 Statistical Analyses	32
3.3.1 Variability Profile	32
3.3.2 Principal Components Analysis (PCA)	33
3.3.3 Cluster Analysis	34
CHAPTER IV: RESULTS	36
4.1 Variability Profile	36
4.2 Principal Components Analysis (PCA)	36
4.2.1 Scores Plot	38
4.2.2 Loadings	38
4.2.3 Scores	42
4.2.4 Bi-Plot	42
4.3 Cluster Analysis	45
4.3.1 Genetic Distance	49
CHAPTER V: DISCUSSION	50
5.1 PCA	50

UNIVERSITI SAINS ISLAM MALAYSIA
ISLAMIC SCIENCE UNIVERSITY OF MALAYSIA

5.1.1 Scores Plot 51

5.1.2 Loadings 51

5.1.3 Scores 53

5.1.4 Bi-Plot 54

5.2 Cluster Analysis 54

5.2.1 Genetic Distance 56

CHAPTER VI: CONCLUSIONS AND RECOMMENDATIONS 57

REFERENCES 59

APPENDIX 65

UNIVERSITI SAINS ISLAM MALAYSIA
ISLAMIC SCIENCE UNIVERSITY OF MALAYSIA

LIST OF TABLES

Page

Table 1: Extent of Variation 36

Table 2: Variability profile of Senegal and Gambian oil palm germplasm as analyzed by descriptive statistics 37

Table 3: Principal components analysis for Senegal and Gambia oil palm germplasm based on 46 agronomic traits 39

Table 4: Scores of the 42 Senegal-Gambian oil palm germplasm on the extracted PCs 43

Table 5: Characteristics means of six clusters generated by Ward's cluster analysis based on 46 agronomic traits 48

Table 6: Inter cluster distance as analyzed by proximity matrix of squared Euclidean distance 49

UNIVERSITI SAINS ISLAM MALAYSIA
ISLAMIC SCIENCE UNIVERSITY OF MALAYSIA

LIST OF FIGURES

Page

Figure 1: Scores plot where PCs 1 and 2 are orthogonal to each other 18

Figure 2: Clustering pattern 19

Figure 3: Scores plot of principal component analysis between percentage variance and number of principal components 38

Figure 4: Scattered diagram of 46 oil palm germplasm traits for first two components contributing almost half of the total variability 41

Figure 5: Scattered diagram of 46 oil palm germplasm traits showing correlation to the first two components 41

Figure 6: Two dimensional ordinations of 42 Senegal-Gambian oil palm accessions on principal axes 1 and 2 44

Figure 7: Bi-plot of 46 oil palm agronomic traits and 42 oil palm accessions on PC 1 and PC 2 44

Figure 8: The relationship among the oil palm germplasm reflected by cluster analysis 46

UNIVERSITI SAINS ISLAM MALAYSIA
ISLAMIC SCIENCE UNIVERSITY OF MALAYSIA
جامعة العلوم الإسلامية الماليزية

LIST OF APPENDICES

Page

Appendix 1: Proximity Matrix of Squared Euclidean Distance 66

UNIVERSITI SAINS ISLAM MALAYSIA
ISLAMIC SCIENCE UNIVERSITY OF MALAYSIA

ABBREVIATION

ABW	average bunch weight
BNO	bunch number
BWT	bunch weight
BDM	bunch dry matter
BI	bunch index
C12:0	lauric acid
C14:0	myristic acid
C16:0	palmitic acid
C16:1	palmitoleic acid
C18:0	stearic acid
C18:1	oleic acid
C18:2	linoleic acid
C18:3	linolenic acid
C20:0	arichidic acid
DIAM	diameter of trunk
e	radiation conversion efficiency
f	fractional interception of radiation
F/B	fruit/bunch
FFB	fresh fruit bunch
FP	frond production
GAM	Gambia
HT	height
IV	iodine value
K/B	kernel/bunch
K/F	kernel/fruit
KY	kernel yield
LA	leaf area

UNIVERSITI SAINS ISLAM MALAYSIA
ISLAMIC SCIENCE UNIVERSITY OF MALAYSIA

LAI	leaf area index
LL	leaflet length
LN	number of leaflet
LW	leaf width
MARDI	Malaysian agricultural research and development institute
M/F	mesocarp/fruit
MFW	mean fruit weight
MNW	mean nut weight
MPOB	Malaysian palm oil board
NAR	net area ratio
NIFOR	Nigerian institute for oil palm research
O/B	oil/bunch
O/DM	oil/dry mass
O/WM	oil/wet mass
OY	oil yield
P/B	parthenocarpic fruit/bunch
PCA	principal component analysis
PC(s)	principal component(s)
PCS	petiole cross section
PORIM	palm oil research institute of Malaysia
RL	rachis length
RRS	reciprocal recurrent selection
SEN	Senegal
S/F	shell/fruit
TEP	total economic product
TDM	total dry mass
UPGMA	unpair weighted group method using arithmetic averages
VDM	vegetative dry mass
WHCA	Ward's hierarchical clustering analysis

UNIVERSITI SAINS ISLAM MALAYSIA
ISLAMIC SCIENCE UNIVERSITY OF MALAYSIA

CHAPTER I

INTRODUCTION

Oil palm (*Elaies guineensis* Jacq.) is a diploid monocotyledon belonging to the family Arecaceae. The crop produces two types of oil, namely palm oil and palm kernel oil. Palm oil is one of the world's most traded vegetable oils in the international market and most widely consumed edible oil (Corley & Tinker, 2003; Rajanaidu, 1994). 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 (Jalani, 1998).

The majority of oil palm plantation and palm oil production is provided by Indonesia and Malaysia, as both contribute about 44 % and 41.5 % respectively to the world palm oil production (Oil World, 2008). 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 in 1848 and afterwards in Malaysia (Corley & Tinker, 2003). 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 (Rajanaidu et al., 1999).

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 programmes (Mandal, 2008). 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 enhanced understanding of the structure, predominantly of large germplasm collections (Martinez-Calvo et al., 2008). In this view, chemometrics or multivariate statistical methods particularly the principal component and cluster analyses have gained wide recognition in the evaluation of germplasm materials of many species (Muhammad et al., 2013; Ajmal et al., 2013; Doumbia, 2013; Lohani et al., 2012; Deyong, 2011; Lacis et al., 2010 and Bozokalfa et al., 2009).

Chemometric is the science of relating measurements made on a chemical system or process to the complex state of the system via multivariate statistical methods (Martens & Naes, 1993). One of the main advantages of chemometric method 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 (Martens & Naes, 1993). The main purpose of the method is to reduce dimensions of complex multivariate data and to simplify data interpretation by finding new orthogonal variables, principal components (PCs), describing the variance in data. Cluster analysis is used in the categorization of germplasm materials into groups based on similarity or dissimilarity (Arroyo, 1993).

1.1 Statement Problem

There is a need to introduce oil palm germplasm as the genetic base of oil palm industry in Malaysia. The oil palm in Malaysia mostly originating from the four palms planted in Bogor in 1848. The germplasm introduced needs to be evaluated and characterized.

1.2 Aim

To study variation in oil palm germplasm by using chemometric analysis [principal component analysis (PCA) and clustering].

1.3 Research Questions

i. Is there variation in the oil palm germplasm?

ii. What are the traits contributing to the variation?

iii. What is the relationship among accessions in the germplasm?

1.4 Objectives

The objectives of the study were:

i. To determine the level of variation in the oil palm germplasm; and

ii. To identify and classify the groups of accessions with different genetic diversity based on quantitative traits.

UNIVERSITI SAINS ISLAM MALAYSIA
ISLAMIC SCIENCE UNIVERSITY OF MALAYSIA
جامعة العلوم الإسلامية الماليزية
USIA

CHAPTER II

LITERATURE REVIEW

2.1 HISTORY OF THE GENUS *Elaeis*

The genus *Elaeis* is derived from a Greek word elaion, which means oil. The oil palm is today grown commercially in South-east Asia, Equatorial America, Africa and South Pacific. The high and increasing yields of the oil palm have led to a rapidly expanding world industry, now based in the tropical areas of Asia, Africa and America (Rajanaidu & Jalani, 1994). Its origin is believed to have been in Africa, but the most productive parts of the industry at present are in Malaysia and Indonesia, which provide most of the oil entering international trade (Oil World, 2012). As a commercial crop, its history is rather short, dating back to 1807 on the West African coast where its cultivation was commenced. It came to the East via the island of Mauritius in 1848, to Indonesia where four seedlings were planted and from there to Singapore Botanical Gardens about 1870 and only then into Malaya in 1875 (Jalani, 2012). The Family Palmae or Arecaceaae, where the genus *Elaeis* belongs is considered as old as any other family of flowering plants with fossils discovered in Cretaceous rocks dating back some 120 million years (Purseglove, 1972). Many plant taxonomists believe it to be the first monocotyledon to have branched out from primitive dicotyledonous stock and as such, the progenitor of all monocotyledons.

The world production of oil palm products has always been impossible to assess accurately owing to the quantities of produce that are not recorded, because they are produced in groves, smallholder plots and farms, and used for the farmer's domestic

purposes or sold locally (Oil World, 2012). Estimates suggest that world-wide production rose from 2.2 million tonnes of palm oil and 1.2 million tonnes of kernels in 1972 to 21 million tonnes of oil, 6 million tonnes of kernels and 2.6 million tonnes of kernel oil in 2000. Most of this increase can be attributed to Malaysia and Indonesia, and to some smaller Asian producers. The production of palm oil has now overtaken that of other vegetable oils, apart from soybean oil. The oil extracted from the oil palm is of two types which are palm oil and palm kernel, of which 90 % is used for food purposes and 10 % for non-food purposes (Jalani, 2012).

2.1.1 Evolution of Genus *Elaeis*

The genus *Elaeis* which belongs to subfamily *Cocoideae* is believed to have originated either in Africa or America and is one of the 240 genera of the family *Arecaceae.* It cannot be viewed in isolation from the other genera because of high degree of homogeneity among the chromosomes of the palms (Dransfield et al., 2005). The first species of the genus, *E. guineensis*, known as the African oil palm was described by Jacquin in 1763. The second species is *E. oleifera* known as the South American oil palm initially described as *E. melanococca* and applied by Gaertner in 1788 to be a form of *E. guineensis*. This species is distinguished from *E. guineensis* by its trunk. Purseglove (1972) however doubted the classification as *E. oleifera* hybridizes easily with *E. guineensis* and the degree of divergence between the two species rather justifies separation on a specific level rather than generic level. The third species was previously known as *Barcella odora*, but was named *Elaeis odora* by Wessels. Boer (1965); it is not cultivated, and little is known about it.

However, molecular markers indicated that inclusion of *E. odora* within the genus *Elaeis* is justified (Barcelos *et al.*, 2002): the genetic distance between *E. odora* and

the other o species of *Elaeis* was similar to the distance between the latter, and less than the distance from *Cocos nucifera*, another member of the Cocosoideae subfamily. The fourth species, *E. madagascariensis,* was described by Beccari (1941) but is still controversial. Some taxonomists believe that these species is a variant of *E. guineensis* which was introduced around the 10th century when Africa influence entered Madagascar (Purseglove, 1972). The species is distinguished from *E.guineensis* by its male flower in which the fused filaments of the staminal tube are shorter and the anthers erect, instead of spreading at anthesis while the fruits are smaller, rounded and surrounded by larger bracts. Uhl & Dransfield (1987) recognized only two species, namely *E. guineensis* and *E. oleifera*, the third species as belonging to *Barcella odora* (Trail) Trail ex Drude and the fourth species as a variant of *E. guineensis* to Madagascar as mentioned above. For the purpose of this research, *E.guineensis* will only be extensively focused on.

2.1.2 Morphology and Biology of *E. guineensis* (African Palm oil)

Elaeis guineensis is a large, pinnate-leaved palm having a solitary columnar stem with short internodes. There are short spines on the leaf petiole and within the fruit bunch. The separate upper and lower ranks of leaflets on the rachis give the palm a characteristic untidy appearance. The palm is normally monoecious with male or female, but sometimes mixed, inflorescences developing in the axils of the leaves. The fruits are borne on a large, compact bunch. The fruit pulp, which provides palm oil, surrounds a nut, the shell of which encloses the palm kernel.

The early descriptions of the oil palm are listed in Hartley (1988). The only one of more than historical interest is the botanical description by Jacquin (1763). He described the palm from material from Martinique (to where it must have been

introduced); his description was detailed, but he described the flowers as either female or *hermaphroditi steriles* and seemed unaware that flowers of the two sexes were in separate inflorescences. The production of male and female inflorescences was first recorded by Miller in his *Gardener's dictionary* (London, 1768).

Before the end of the eighteenth century Gaertner (*De fructibus et seminibus plantarum*, Stuttgart, 1788) gave a more detailed description of the flower parts, recording that the male and female flowers are on separate inflorescences. Most of the early attempts at classification of varieties were unsatisfactory, as they were based on very limited acquaintance with the palm, and no knowledge of the inheritance of the characters described. However, it is the first description by Preuss (1902) of the *lisombe* palm, a name used in Congo, Cameroon and Nigeria for the thin-shelled *tenera* fruit form and still employed in quite recent times.

Janssens (1927) and Smith (1935) provided the first simple classifications which, in their essentials, have stood the test of time. Although nothing was known of the inheritance of the characters described, Janssens recognised that the fruit forms *dura* and *tenera*, distinguished by the thickness of shell, could be found in fruit types of different external appearance. Thus, both the common fruit type *nigrescens* and the green-fruited *virescens* were divided by Janssens into three forms, *dura*, *tenera* and *pisifera*.

2.1.3 Infraspecific Classification of *E. guineensis*

The taxonomy of *E. guineensis* variant forms has not yet been clarified. Several authors have described them as cultivars, varieties, subspecies or races. Purseglove (1972) believed that cultivars in the real sense do not occur in the species. He

therefore produced an infraspecific classification based on fruit characters, *viz., dura* (DD), *tenera* (Dd) and *pisifera* (dd), without committing them to specific taxonomic categories. Other author such as Whitemore (1979) debated the various forms *idolatrica*, *tenera*, *deli dura* and *pisifera* as varieties. As categories such as subspecies, variety or form may be inappropriate in the infraspecific classification of the species, the term race is proposed. Race is defined as a permanent variety or a microspecies. This term may prove appropriate because, in the genetic sense, races of a species are capable of exchanging genes and this has been amply demonstrated by hybridization between the races of *E. guineensis*.

Brief Descriptions of the various common races of *E. guineensis* (Willy, 2010).

- Race *dura* (DD): endocarp 2mm-8mm thick, comprising 25 % - 55 % of fruit weight; medium mesocarp content of 35 % - 55 % by weight; kernel tends to be large comprising 7 % - 20 % of fruit weight. The *Deli dura* is an ecological variant of *dura*, differing from latter in only minor differences such as kernel content (4 % - 10 %), mesocarp content (up to 65 %) and as such larger fruit size. The *macrocarya* form with endocarp 6mm-8mm thickness is an extreme *dura* race.
- Race *tenera* (Dd): endocarp 0.5mm-3mm thick, comprising 1 % - 32 % of fruit weight; medium to high mesocarp content of 60 % - 95 %; fibre ring darker in colour. The dividing line between *dura* and *tenera* is sometimes rather distinct as shown by the presence of a fibre ring in the mesocarp of the latter which is diagnostic; kernel usually smaller, comprising 3 % - 15 % of fruit. The oil content is higher at 24 % - 32 %.

UNIVERSITI SAINS ISLAM MALAYSIA
ISLAMIC SCIENCE UNIVERSITY OF MALAYSIA

- Race *pisifera* (dd): no endocarp, with small pea-like kernel in fertile fruits, but predominantly sterile with fruits rotting prematurely. It has a much higher sex ratio than *dura* and *tenera* and a significantly higher number of spikelets; even when fertile, the fruit to bunch ratio is low; infertile palms show strong vegetative growth.

2.1.4 Ecology and Morphology of *E. guineensis*

In its natural habitat in West Africa, *E. guineensis* often occur in association with *Raphia* or, if alone, in fresh water swamps (Chevalier, 1943). It is absent from undistributed tropical rain forests, where there is competition from the forest flora, absence of light due to the forest canopy and excessive moisture. It cannot survive or regenerate in high secondary forests for the same reasons. It may tolerate temporary flooding, if the water is not stagnant. Under semi-wild conditions, *E. guineensis* groves owe their presence to alteration of the natural vegetation by man for their development. In southern Nigeria, with its high population pressure, oil palm reaches its highest development and has given rise to palm groves. In both wild and semi-wild conditions, the yields are very low compared to plantation conditions, where individual plants are properly spaced and adequately cared for (Corley &Tinker, 2003).

The biggest area of semi-wild *E. guineensis* groves in the lowland regions of Western and Central Africa lies between 10° N and 10°S, which have a marked dry season lasting up to four to six months. The genus is ecologically suited to such conditions, although physiologically, it lowers yields. The present knowledge indicates that under plantation conditions, *Elaeis* cultivation can be extended from its natural habitat of 10°N and 10°S of Equator to 23°N and 23°S, i.e. between the tropics of Cancer and

Capricon. This has been brought about by man's deliberate planting of oil palm as it is an important component of world's vegetable oil supply (Latiff, 2000; Willy 2010).

It is generally known that *Elaeis* thrives well between a mean minimum temperature of 20°C - 23°C and a mean maximum temperature of 28°C - 32°C which normally occur in tropical countries. If the temperature falls below this, particularly at night temperature of below 19°C, bunch development is affected and the yield reduced (Latiff, 2002). There is also evidence to show that a temperature below 15°C stops growth in young seedlings. In Malaysia, oil palm can be grown on a wide range of soils. It is found that adequate soil moisture is more important than nutrient supply, which can be artificially supplied. In Peninsular Malaysia, the best areas are the coastal alluvial clay; Sabah's the riverine and coastal alluviums, and the soils of volcanic origin.

Morphology

- The oil palm has a typical adventitious root system. When the seed germinates, the radicle emerges through the germ-pore. The radicle continues to grow and is joined when the seed becomes detached by the primary roots. The short-lived radicle is replaced by adventitious roots emerging from the periphery of the radicle-hypocotyl junction. Primary roots emerge from the base of the swollen part. The primary roots descend deeply from the base of the trunk and spreads horizontally at varying depths in soil as the palm develops. They are up to 6mm-10mm in diameter and up to 20m in length. The secondary roots are produced from the pericycles of the primary roots amounting to 2mm-4mm in diameter ascending to the surface of the soil (Purseglove, 1972; Corley &Tinker, 2003).

- A trunk of an oil palm is not visible until the palm is three years or so old when the apex reaches its full diameter inform of an inverted cone. The base of the trunk is about 60 cm in diameter and the trunk is about 40 cm in diameter. The rate of trunk extension is about 25 cm - 50 cm per year. All mature palms have a solitary columnar stem with persistent frond bases. The stem of *E. guineensis* is stout and despite the crowded steeply ascending frond petioles bases, and then ultimately smooth. Frond bases persistently adhere to the stem for at least 12 years and then fall away except for a few near the crown. The stem measures between 22 cm to 75 cm in diameter (Latiff, 2002).
- The frond consists of leaflets, each with a lamina and midrib, a central rachis to which the leaflets are attached, a petiole and a frond sheath. Only remnants of the frond sheath are visible externally. In a developing frond, the sheath is tubular and completely encloses but as it extends after growth of the sheath has stopped, the fibrous sheath splits and breaks up, leaving a row of spines along both edges of the petiole which are bases of fibrous sheath. The leaflets are long, ranging from 55 cm to 65 cm and often 100 cm. it is narrow with a width from 2.5 cm to 4.0 cm and a midrib with a number of parallel veins on the lamina. The cuticle is thick and has a very high resistance to diffusion of water vapour. Stomata are situated on the abaxial surface of the leaflets only. The leaflets are arranged in two planes (Latiff, 2002; Willy, 2010).
- The oil palm produces either a male or female or at certain stages a hermaphrodite inflorescence in each of the frond axil during the mature stage. An average female inflorescence may have more than 100 spikelets with over 4000 floral buds. The spikelets develop acropetally on the inflorescence. They are about 30 cm long with 12 - 30 flowers on central spikelets and less than 12

flowers on the lower and upper spikelets. Averagely, there are 700 - 1200 flowers per spikelet. Therefore, the total number of flowers is about 126000 and the number of pollen grains has been estimated to be in excess of 900 million with weight of 40 g per inflorescence. The presence of empty frond axils is an indication of abortion. The hermaphrodite inflorescence in *E. guineensis* is similar to that of *E. oleifera*. It has about 200 spikelets of varying lengths from 5 cm – 15 cm (Willy, 2010).

- The oil palm gives the highest yield of oil per unit area of any crop and produces two distinct oils, palm oil and palm kernel oil, both of which are extremely important in world trade. The time from flowering to harvesting of ripe fruits is five to six months. The fruits are borne on spikelets which are spirally arranged to form a compact bunch. The fruits are drupes of about 10 kg – 90 kg in weight. The pericarp comprises of three layers; exocarp (external layer), mesocarp (outer layer) and endocarp (internal layer). The exocarp protects the fruit; the mesocarp contains the oil and the endocarp is a hard shell containing the kernel which contains the kernel oil and the endosperm (Latiff, 2002; Willy, 2010).
- The seed is a nut which remains after the soft oily mesocarp has been removed usually by retting. It consists of an endocarp and possibly two or three kernels. The nut size varies greatly and is dependent both on the thickness of shell and size of kernel. It may be 2 cm - 3 cm in length and average 4g in weight. Deli *dura* and African nuts are larger, weighing up to 12 g, and the African *tenera* nuts are usually 2 cm or less in length and average 2 g (Corley &Tinker, 2003).

UNIVERSITI SAINS ISLAM MALAYSIA
ISLAMIC SCIENCE UNIVERSITY OF MALAYSIA

2.2 PLANT GENETIC RESOURCES

Plant genetic resources include the variation in crop plant genetic material that is accessible for present and potential future exploitation (FAO, 1996). The variation contains diversity which can be found at the nucleotide sequences levels, genotypes and alleles, and is required for the improvement of new cultivated varieties along with contributing to the flexibility of existing varieties (Hammer et al., 2003). Plant genetic resources are significant basis of genetic diversity and contribute to food security. Plant breeding is founded on the exploitation of genetic diversity within and between crop species and varieties (FAO, 1996).

2.2.1 Genetic Diversity

Genetic diversity has been defined as the germplasm of plants, animals or other organisms containing useful characters of actual or potential values, especially when these characters provide the variation in genes and genotypes between and within species or populations (Cromwell et al., 1999). It is the diversity that enables species to adapt to changing environments and provides an insurance against unknown future needs or conditions, thereby contributing to stability of farming systems at the local, national and global level. Genetic diversity comprises the total genetic variation present in a population or species. It is the differences within and between species or varieties in genes, alleles and genotypes, caused by mutation and recombination. Genetic diversity is the basis of selection in crop plants, it is essential for the development of new varieties by using novel combinations and traits (Preston, 2011).

In the field, genetic diversity among and between individuals and varieties is essential for resistance to pests and diseases, as well as tolerance and adaptation to climatic

conditions and changing climate (Preston, 2011). According to Newbury and Ford-Lloyd (1997), maintenance of the range and magnitude of genetic diversity present within a taxon is a primary aim of plant conservation. Therefore, to facilitate the conservation of genetic variation for present and future use, and to establish a baseline, the diversity of plant genetic resources, such as those held at MPOB needs to be characterized.

2.2.2 Characterization of Genetic Diversity

The level of genetic diversity present in populations is influenced by many factors including life form, breeding system, seed dispersal and geographic range. The overall result of this is a higher level of genetic diversity within populations and lower level of differentiation between populations in outbreeding species; a lower level of genetic diversity within populations and higher level of genetic diversity between population differentiations in inbreeding species (Hamrick & Godt, 1996). Genetic patterns of diversity may be variable across time such as in species due to changes in agriculture, and such as in ex-situ collections (due to genetic drift, cross pollination) and selective effects during regeneration, which can result in changes in genetic diversity and allele frequencies (Negra & Tiranti, 2009).

Genetic diversity has been explored using, predominantly, three marker types; morphology, proteins and DNA based methods. Before the advent of modern genetic technology, morphological markers were the classical method for characterization or estimation of genetic diversity. These comprised a diversity of traits and measurements that were recorded at all stages of plant development. Using morphological markers confer many advantages and morphological studies are often the first step in species studies and plant genetic resource activities serving to inform

molecular studies as to where diversity may exist (Karp et al., 1997). Advantages comprise the low cost and level of technology required, and the relation of markers to traits of agronomic importance (Newbury & Ford-Lloyd, 1997).

However, there remain limits to the usefulness or applicability of morphological markers, many of which are met by molecular methods. These include: the limited number of informative characters available, some of which may show little variation over material. The quantitative nature of their inheritance (being jointly influenced by genetics and environmental conditions of growth, because of this, some traits cannot be reliably isolated); related to this is the effect of environment that may mask the genetic coordinate and therefore influence genetic diversity estimates based on morphology (Spooner et al., 2005).

In the past, determination of the genetic structure of germplasm collections has mainly been done using passport data (van Hintem, 2000) or multivariate statistical methods such as cluster analysis, principal component analysis, multidimensional scaling; usually based on agronomic data (Franco et al., 2001).

2.3 CHEMOMETRICS

The development of the discipline chemometrics is strongly related to the use of computers in chemistry. Chemometrics is the chemical discipline that uses mathematical and statistical methods to design or select optimal measurement procedures and experiments, and to provide maximum chemical information by analyzing chemical data (Matthias, 2007). Chemometrics is the use of mathematical and statistical methods for selecting optimal experiments and extracting maximum amount of information when analyzing multivariate data. Chemometrics can also be

called multivariate analysis (Matthias, 2007). This is because the statistical method of classification is usually by multivariate methods which include Principal Component Analysis (PCA), Cluster analysis and Discriminate analysis (Oyelola, 2004). They are all under exploratory statistical analysis. Wold and Sjostrom (1998) have reviewed that one of the successful area in chemometrics was pattern recognition method which is applied in both industry and academia.

Chemical pattern recognition method often used in chemometrics to seek the pattern from the multivariate data. In addition, Gonzalez (2012) has discussed that pattern recognition was part of Artificial Intelligence field that focused on finding the similarities and variation from large data sets in which brought about natural classification and grouping. In most cases, the raw data from the chemistry research were in form of chromatographic or spectroscopic. Hence, the chemist has dealt with chemometrics analysis to extract the maximum of useful information, thus, highlighted the differences in the chromatographic or spectroscopic results of different variables. As a result, exploratory data analysis (EDA) gave simple earlier visual idea of the main relationship between the objects and the variables (Brereton, 2003).

2.3.1 Principal Components Analysis (PCA)

PCA is a widely used mathematical tool for high dimension data analysis; that provides a guideline for how to reduce a complex dataset to one of lower dimensionality, to reveal any hidden, simplified structures that may underlie it (Jolliffe, 2002). Although PCA is a powerful tool capable of reducing dimensions and revealing relationships among data items, it has been traditionally viewed as a "blackbox" approach that is difficult to grasp for many of its users because of its coordinate transformation from original data space into Eigen space. PCA is a method

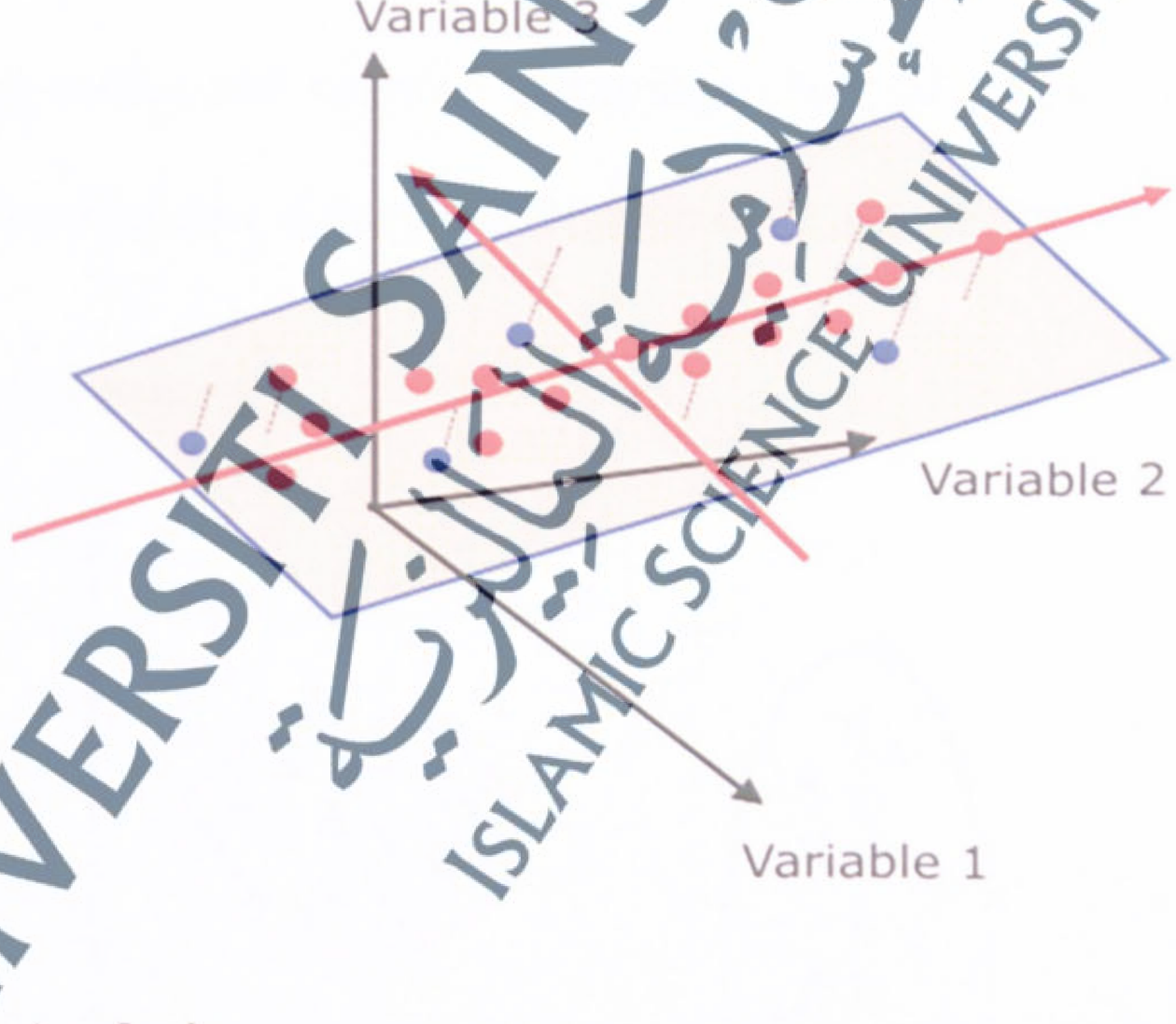
that projects a dataset to a new coordinate system by determining the eigenvectors and eigenvalues of a matrix. It involves a calculation of a covariance matrix of a dataset to minimize the redundancy and maximize the variance. Mathematically, PCA is defined as an orthogonal linear transformation and assumes all basis vectors are an orthonomal matrix (Jolliffe, 2002). PCA is concerned with finding the variances and coefficients of a dataset by finding the eigenvalues and eigenvectors.

PCA is one of the most useful statistical tools for screening multivariate data with significantly high correlations. Information from PCA may assist the plant breeder to identify limited traits for using in hybridization and selection programs (Doumbia et al., 2013). The central idea of principal component analysis (PCA) is to reduce the dimensionality of a data set consisting of a large number of interrelated variables, while retaining as much as possible of the variation present in the data set. This is achieved by transforming to a new set of variables, the principal components (PCs), which are uncorrelated, and which are ordered so that the first few retain most of the variation present in all of the original variables. PCA tool is used to interpret the variables that described the differences between samples, thus, identifies which variables contributed most to an observed difference or which variables were correlated. In short, PCA is the most basic workhorse of multivariate data analysis that acts as fundamental technique for other methods in "The Unscrambler" (CAMO, 2014).

Basically, the principle of PCA aims to find the direction in multidimensional space along which the distance between the data points is the largest thereby indicating the maximum variance in data set. In other words, it constructed as linear combinations of the initial variables that contributed most in making the samples different from each

other, thus minimizing the squared distance from the line to each point of data set. Data point was the representation of each sample in a data table, thus, the location was determined by the cell values of the corresponding row in the table. Thus, each variable act as coordinate axis in multidimensional space. PCs are computed iteratively where the first PC contained largest information or most explained variance while the subsequence PC contained less explained variance than previous one. Based on the geometric concept, a new set of coordinate axis were formed as shown in Figure 1 below where PCs were orthogonal to each other. Thus, the user can focus on the first few PCs in the interpretation which contained larger information than the subsequence ones (Esbensen *et al*., 2002; CAMO, 2014).

Figure 1: Score plots where PCs 1 and 2 are orthogonal to each other (CAMO, 2014)



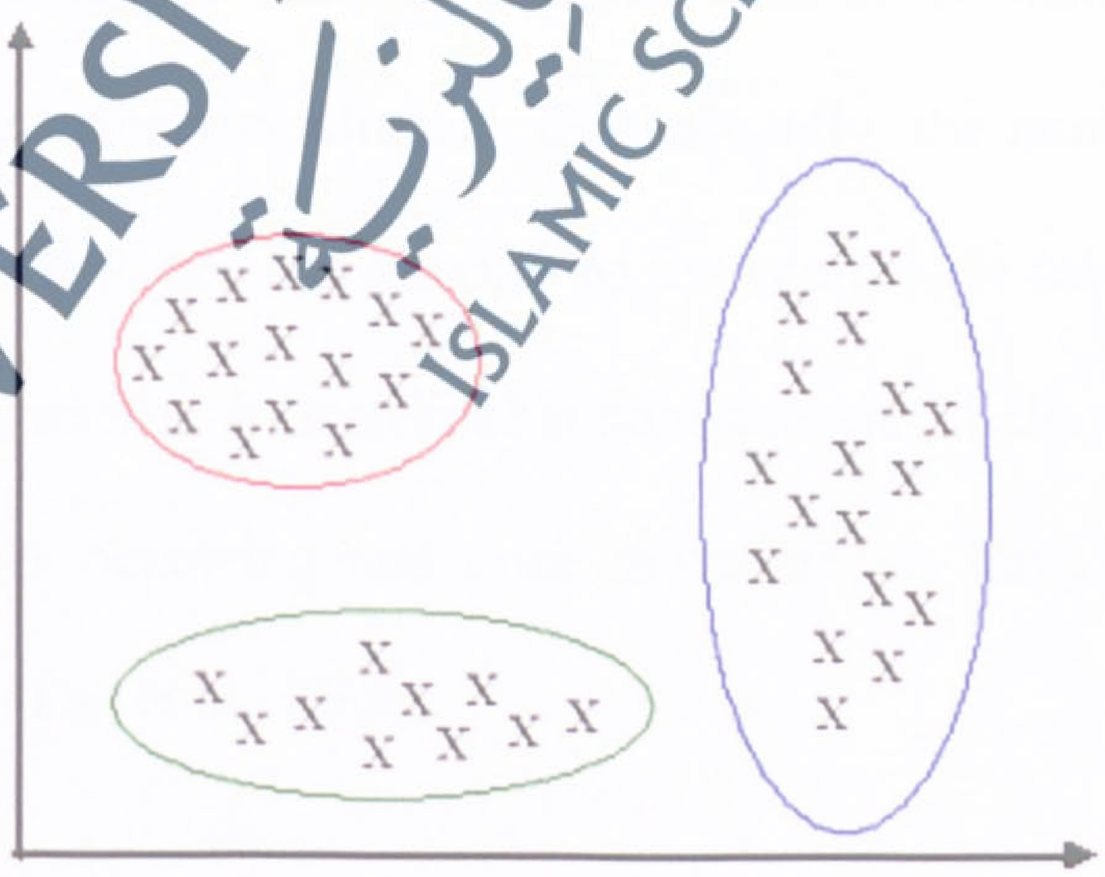
2.3.2 Clustering Analysis

Clustering deals with finding a structure in a collection of unlabeled data. A loose definition of clustering could be "the process of organizing objects into groups whose members are similar in some way". A cluster is therefore a collection of objects, which are "similar" between them and are "dissimilar" to the objects belonging to

other clusters (Manpreet et al., 2008). Thereby, clustering is the grouping of data items based on their similarity. Cluster analysis is classified as unsupervised pattern recognition due to the discovery of the data structures without providing any explanation as well as the pre-information about the data not being used in the classification (CAMO, 2014).

The goal of clustering is to determine the intrinsic grouping in a set of unlabeled data. But how to decide what constitutes a good clustering? It can be shown that there is no absolute "best" criterion, which would be independent of the final aim of the clustering. Consequently, it is the user, which must supply this criterion, in such a way that the result of the clustering will suit their needs. For instance, we could be interested in finding representatives for homogeneous groups (data reduction), in finding "natural clusters" and describe their unknown properties ("natural" data types), in finding useful and suitable groupings ("useful" data classes) or in finding unusual data objects (outlier detection) (Manpreet et al., 2008).

Figure 2: Clustering pattern



Clustering can be in the form of distance-based clustering and conceptual clustering.

Distance – based cluster is when two or more objects that belong to the same cluster

are "close" according to a given distance. On the other hand, conceptual clustering occurs when two or more objects that belong to the same cluster are defined by common concept. In other words, objects are grouped according to their fit to descriptive concepts not according to simple similarity measures.

2.3.2.1 Types of Clustering

The Unscrambler software offers two main groups of cluster analysis which are classified as non-hierarchical clustering (*k*-means and *k*-medians) and hierarchical cluster analysis (HCA).

K-means Clustering

The target of the clustering is to minimize the variability of the samples within the clusters along with maximizing the variability between clusters formed starting with *k* random clusters (StatSoft, 2013). The theory of *k*-means clustering algorithm provided by the Unscrambler starts with the randomized set of samples distribution between the defined *k* random clusters. Every defined clusters centroid is calculated using averages (*k*-means) method, then, the distance between each of the samples to its centroid is measured within the respective clusters. Consequently, the number of clusters *k* with the shortest distance between the samples to the centroid is selected by moving a set of samples to that particular *k* cluster. This nonhierarchical algorithm is repeated until there are no changes occurring and once the centroids have identified the natural clustering of points (Tan et al., 2006).

K- median clustering

In statistics and data mining, k-medians clustering is a cluster analysis algorithm. It is a variation of k-means clustering where instead of calculating the mean for each

cluster to determine its centroid, one instead calculates the median. This has the effect of minimizing error over all clusters with respect to the 1-norm distance metric, as opposed to the square of the 2-norm distance metric (which k-means does.). K-median is used when one wishes to minimize the total 1-norm distance from each point to its nearest cluster center (Jain & Dubes, 1988; Bradley et al., 1997).

Hierarchical Clustering Analysis (HCA)

HCA used different linkage methods along with specific distance measure to create clusters. The chosen of both factors need to be appropriately selected based on the application domain as well as base on the real-world interpretation because they can affect the grouping results. Consequently, a dendrogram is created to portray the results of clusters arrangement produced yet formed based on the meet of triangle inequality of metric determined from the distance between samples (CAMO, 2014). Basically, HCA begins by treating each sample within its class, then, proceeded by combining the samples into another cluster based on their similarity until the clusters emerged to form one larger cluster (CAMO, 2014).

As discussed earlier, the distance measure between samples is important to be determined for the first part in relating the samples into the same group based on its closest distance. Hence, the Unscrambler offers several options for the distance measures between samples such as Euclidean distance, squared Euclidean distance, city-block distance and Chebyshev distance. Among these methods, Euclidean distance is the most usually chosen in measurement of the distance between two samples of the data sets that are suitably normalized by taken into consideration the different between two samples directly based on the magnitude of changes in the sample levels. The squared Euclidean distance intended to normalize the data by

measuring the similarity between clusters where some variable may control the distance between groups (Hoon *et al.*, 2013 and CAMO, 2014).

2.4 PREVIOUS STUDIES

Principal component analysis and cluster analysis have been widely used in various studies for delineating the variability in large group of genotypes of many crop species, some of which are outlined below:

Shivani and Sreelakshmi (2014) assessed the genetic diversity in indigenous germplasm lines of safflower using multivariate tools. Principal component analysis revealed that 99% of the total diversity was explained on the basis of the first three principal components. The distribution of accessions within eight well defined clusters was visible as well with no apparent relationship with the geographical origin.

In evaluating the variation in seed vigour characters of West African rice (*Oryza sativa L.*) genotypes, seeds of 24 West African rice genotypes were selected based on seed vigour traits in the laboratory and field in two cropping seasons at the research farm of Federal University of Agriculture, Abeokuta, Nigeria. The analysis was subjected to principal component analysis and clustering analysis. The first three axes of the PCs across the two seasons captured 86.34% of the total variation. Cluster analysis classified these genotypes into four distinct groups based on germination and emergence percentages. Those characters identified by PCA could be included in the crop improvement programme for improved seed quality within West African low land rice germplasm (Adebisi et al., 2013).

Genetic diversity based on cluster and principal component analyses for yield and quality attributes in ginger was studied by Ravishanker et al. (2013). Through cluster

analysis, 25 genotypes were grouped into five main clusters while first six PCs with eigenvalues greater than one accounted for 76.19% of the total variability.

The comparative study of cowpea germplasms diversity from Ghana and Mali using morphological characteristics was carried out by Doumbia et al. (2013). Results from PCA indicated that first PC explained a total variation of 72.05% while UPGMA showed that at 0.97 level of similarity, almost all the 94 accessions studied were distinct from each other while at 0.95 levels, they were similar to each other.

Muhammad et al. (2013) utilized PCA and cluster analysis to study the genetic divergence in indigenous spinach. Results obtained shows that the cumulative contribution of the first three PCs reflected by PCA was 58.4 % while cluster analysis grouped the spinach genotypes into four major clusters regardless of collection origin.

In the research of WeldeMicheal et al. (2013), forty nine coffee germplasm (*Coffea arabica L.*) accessions were collected from Gomma Wereda in other to estimate information on genetic diversity. The experiment was conducted in simple lattice design with two replications during cropping season by superimposing on six years old coffee trees and grown under uniform coffee shade tree (*Sesbania sesban*) conditions. Data from 26 quantitative characters were recorded. The analysis of variance showed a significant variation among the accessions for all morphological traits. This shows the existence of variability among the tested materials. Cluster analysis grouped the 49 germplasm accessions into five clusters which make the accessions to be moderately divergent. The distances between these clusters are highly significant ($P<0.01$) which suggested a suitable hybridization program is possible. Principal component analysis showed a variation in the first two principal components which is explained by the larger share of the observed variation.

Ajmal et al. (2013) used multivariate analysis to study the genetic relationships among wheat germplasm comprising of 50 genotypes. The first three PCs with eigenvalues greater than one contributed to 70.59% of the variability amongst genotypes while cluster analysis set apart the 50 genotypes into 5 clusters based on Ward's method.

Principal component analysis was done by Lohani et al. (2012) to estimate the variability in germplasm of potato. The PCA of 54 potato genotypes based on correlation matrix of agronomic and quality traits showed that first 11 components explained 96.25% variation. Non-hierarchical Euclidean Cluster analysis based on PCA grouped the 54 potato genotypes into seven non-overlapping clusters.

Darvishzadeh et al. (2012) studied on the genetic diversity in population of Iranian Ajowan (*carum copticum L.*) based on agronomical and morphological characteristics. Ten populations were collected from different regions and evaluated in a completely randomized design with eight to ten replications. Among the characteristics studied, a high coefficient of variation was observed for the number of seeds (197.58), plant yield (57.56) and shoot dry matter (56.28). Cluster analysis using Ward's method classified ten populations of Ajowan into four groups. PCA using 18 agronomical and morphological characteristics indicated that the first two principal components with eigenvalue of more than one accounted for 74.5% of the total variance. Both PCA and cluster analysis result were consistent which displayed a considerable diversity of agronomical and morphological is useful in germplasm management.

In order to reveal the underlying association between agronomic traits of spring wheat in eastern region of Qinghai Tibet Plateau, Deyong (2011) used partial correlation analysis, PCA and cluster analysis to study the relationship between yield

grain and agronomic traits. PCA was able to indicate genotypes with high-yield potential and the varieties were grouped into 14 groups through cluster analysis.

Sanni et al. (2010) studied on the variability of 434 accessions of rice germplasm from Cote d'Ivoire and evaluated 14 agro morphological traits in upland condition using experimental design and analyzed with multivariate methods. The result suggested that traits such as plant height, leaf length, tillering ability, number of days to heading and maturity, panicle length and grain size were the principal discriminatory characteristics. The result of PCA indicated that first three PCs explained about 58.41% of the total variation among the 14 characters studied. Seven cluster groups were obtained using the unweighted variable pair group method of the average linkage cluster analysis.

Multivariate statistical analysis was used by Lacis et al. (2010) to determine phonotypical variability and genetic diversity among 40 sour cherry accessions. Both cluster analysis and PCA showed an adequate grouping of accessions according to phenotypical data and known pedigree data. Nine components with eigenvalues larger than 1.0 were extracted which described 80% of the variability of the original traits while cluster analysis identified four main clusters.

Phenotypic diversity for quantitative and qualitative traits in a collection of pepper germplasm from different areas of Turkey was assessed by Bozokalfa et al. (2009). All the accessions were characterized for 67 agro-morphological traits and the morphological traits were subjected to PCA followed by hierarchical agglomerative clustering. The first six PC axes accounted for 54.29% of the variance among the 48 accessions and their lines while seven groups were created based on morphological and agronomic properties.

Peerasak et al. (2009) evaluated genetic variation in cultivated mungbean germplasm to assess the extent and pattern of their diversity. He evaluated 9 qualitative and 21 quantitative traits in 340 diverse cultivated mungbean accessions collected at the world vegetable center, Taiwan. The germplasm displayed a wide range of diversity for most of the traits evaluated. Cluster analysis grouped the germplasm into 5 major and 1 minor cluster. Principal component analysis revealed that the first three PCs explained 74.9% of the total variation. It is recommended that germplasm from west Asia be exploited more in cultivar programs. The results can help mungbean breeders choose the right combinations of parental genotypes carrying the desirable characters.

In the study of genetic variability of some quality traits in *Lathyrus spp*. germpalsm, 66 accessions representing eighteen species of the genus lathyrus were collected from different regions evaluated for variation of quality traits. Cluster analysis of coefficient of variance values for each accession identified the 66 accessions into 8 groups. High variability was attributed both genetic and environmental factors exhibited at both inter-specific and intra-specific levels. The most promising accession for breeding programs was *L. sativus* from Tunisia (Sammour et al., 2007).

UNIVERSITI SAINS ISLAM MALAYSIA
ISLAMIC SCIENCE UNIVERSITY OF MALAYSIA

CHAPTER III

MATERIALS AND METHODS

3.1 Breeding Materials and Site Location

The germplasm used in this study originated from Senegal and Gambia. A random sample of five families from 13 sites originating from Senegal and also Gambia (population; designated as SEN 01, SEN 02, SEN 03, SEN 04, SEN 05, SEN 05, SEN 06, SEN 07, SEN 08, SEN 09, SEN 10, SEN 11, SEN 12, SEN 13 GAM05.02 and GAM05.08) were collected in July-August 1993 by researchers from Malaysian Palm Oil Board (MPOB) from the two countries. The germplasm materials were planted at 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.

3.2 Data Collection

The performance of progenies for yield and oil yield components was assessed based on data on fresh fruit bunch (FFB) yield records, bunch number (BNO) and average bunch weight (ABW), bunch components, vegetative and physiological characters and fatty acid characters.

UNIVERSITI SAINS ISLAM MALAYSIA
ISLAMIC SCIENCE UNIVERSITY OF MALAYSIA

3.2.1 Yield and Yield Components

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. Data on yield and yield components were evaluated from year 2000 – 2007.

Procedure for Yield Recording

Recordings of bunch weight (BWT) and bunch number (BNO) on individual are carried out during the harvesting rounds in oil palm breeding. Fresh bunch weight (FFB) is the sum of the bunch weight (BWT) while BNO is the total of all the bunch counts and average bunch weight (ABWT) is the quotient between FFB and BNO. The yield components were derived as follows:

$$\text{FFB (kg/p/yr)} = \sum_{i=1}^{n} BWT_i$$

$$\text{BNO (bunches/p/yr)} = \sum_{i=1}^{n} BNO_i$$

$$\text{ABWT (kg/p/yr)} = \text{FFB/BNO}$$

Where n is the number of harvesting rounds.

3.2.2 Bunch Analysis

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. (1963).

Procedure for Bunch Analysis

Samples of three to five bunches from each palm were analysed. Each bunch was weighed and fruit bearing spikelets chopped off the stalk. The spikelets were randomly segmented for analysis of the fruit to bunch (F/B) and fruit compositions (FC). To ease picking, the F/B portions were kept for two days and the weight of empty spikelets, fertile and parthenocarpic fruits were recorded. Analysis of the fruit compositions was continued following spikelet sampling on the same day. Fruits were detached and randomly sampled, counted, weighed and scraped. The nuts were crack opened and kernel weighed. A 5g sample of minced mesocarp was oil extracted in hexane for 18-24 hours using soxhlet apparatus. The amount of oil in the sample was calculated on weight difference before and after extraction. The components of the bunch were calculated using the below formulae:

F/B = Fruit to Bunch (%) = [(FFWT + PFWT) / SWT] × [(BWT -STKWT) / BWT] × 100

P/F = Parthenocarpic to Fruit (%) = PFWT / (FFWT + PFWT) × 100

M/F = Mesocarp to Fruit (%) = [(FSWT - FNWT) / FSWT] × 100

MC = Moisture Content of mesocarp(%) =100 -[{(TDMWT-TWT)/(FSWT-FNWT)}×100]

O/DM = Oil to Dry Mesocarp (%) = [(ETMWT - ETFWT) / (ETMWT - ETWT)] × 100

O/WM = Oil to Wet Mesocarp (%) = [(100 - MC) × O/DM] / 100

O/B = Oil to Bunch (%) = (F/B × M/F × O/WM) / 10,000

O/Fi = Oil to Fibre (%) = [(ETMWT -ETFWT) / (ETFWT -ETWT)] × 100

K/F = Kernel to Fruit (%) = (KWT / FSWT) × 100

MNW = Mean Nut Weight (g) = FNWT /NOFNUT

MFW =Mean Fruit Weight (g) = FSWT / NOFNUT

P/B = Parthenocarpic to Bunch (%) = (P/F × F/B) / 100

OPY = Oil per Palm per Year (kg) = (O/B × FFB) / 100

UNIVERSITI SAINS ISLAM MALAYSIA
ISLAMIC SCIENCE UNIVERSITY OF MALAYSIA

KPY = Kernel per Palm per Year (kg) = (K/B × FFB) / 100

Where;

BWT = Bunch weight

SWT = Spikelet weight

PFWT = Parthenocarpic fruit weight

FSWT = Fruit sub sample weight

NOFNUT = Number of fresh nut

TDMWT = Tin + dry mesocarp weight

ETWT = Extraction thimble weight

ETMWT = Extraction thimble + mesocarp weight

STKWT = Stalk weight

FFWT = Fertile fruit weight

ESPKWT = Empty spikelet weight

FNWT = Fresh nut weight

KWT = Kernel weight

TWT = Tin weight

ETFWT = Extraction thimble + fibre weight

FFB = Fresh fruit bunch

3.2.3 Vegetative Measurements and Physiological Characters

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 (1986). Data on the physiological characters were assessed in the year 2007. The various components of the vegetative and physiological parameters were calculated using the below formulae:

FP = Frond production (no/p/yr) = number of fronds produced in one year

PCS = Petiole Cross Sectional Area (cm^2) = petiole depth × petiole width

RL = Rachis length (m) = length from tip of rachis to the first ligule

LL = Leaflet Length = (LL1 + LL2 + ...LL6) / 6

LW = Leaflet Width = (LW1 + LW2 + ...LW6) / 6

LN = Leaflet number (no/p/yr) = (number of fronds on one side of rachis)×2

D = Trunk diameter (cm) = diameter of trunk at one meter from ground

HT = Trunk height (m) = height of trunk of ground to base of frond 41

HTI = Height increment = (HT at year t) / (age at year t - 2)

LA = Leaf Area (m^2) = $[(\Sigma (LL_i \times LW_i) \times LN \times 0.57) / 6] / 10{,}000$

LAR = Leaf Area Ratio = (FP × LA) / VDM

LAI = Leaf Index Ratio = (40 fronds × LA × PD^{+}) / 10,000

LDW = Leaf Dry Weight (kg) = (0.1023 × PCS) + 0.2062

TDW = Trunk Dry weight (kg) = $3.142 \times (D/2)^2 \times (HT / \text{Age in year}) \times 1{,}000 \times 0.17$

FDW = Frond Dry weight (kg) = FP × LDW

FI = Frond Index = LA / LDW

f = Fractional interception = $1 - \exp^{(-0.47) \times (LAI - 0.3)}$

e = Conversion Efficiency (g/MJ) = TDM / (31 × f)

VDM = Vegetative Dry Matter (kg/p/yr) = (FDW + TDW)

BDM = Bunch Dry Matter (kg/p/yr) = 0.53 × FFB

TDM = Total Dry Matter = [(VDM + BDM) × PD^{+}] / 1,000

BI = Bunch Index = BDM / (BDM + VDM)

NAR = Net Assimilation Rate = TDM / (0.52 × LAI)

TOIL = Total Oil (kg/p/yr) = OPY + (KPY x 0.5*)

TEP = Total Economic Product (kg/p/yr) = OPY + (KPY × $0.6^{\#}$)

Where;

$^{+}$PD = Planting Density
*0.5 = Kernel Oil Extraction Rate
$^{\#}$0.6 = Relative Price of Palm Kernel Oil to Price of Palm Oil
FFB = Fresh Fruit Bunch (kg/p/yr)
OPY = Oil per Palm Year (kg)
KPY = Kernel per Palm per Year (kg)

3.2.4 Fatty Acid Traits

Fatty acid traits were pooled between years 2000 – 2007. The fatty acid composition was evaluated using the method proposed by Timms (1978) for routine analysis using gas chromatography.

Procedure for Fatty Acid Analysis

The fatty acid analysis was carried out using MPOB method Timms (1978). Fresh bunches were collected from the palm, weighed and chopped. Spikelets samples were selected one each from apical, middle and basal portion of the bunch. The samples

UNIVERSITI SAINS ISLAM MALAYSIA
ISLAMIC SCIENCE UNIVERSITY OF MALAYSIA

were sterilized and fruits from the sterilized spikelets separated. The mesocarps were separated and oven dried at 105 °C for a minimum of 3 hours. The dried mesocarp were minced and dried. After mincing, the dried mesocarp was diluted with 300ml hexane and filtered with a spoonful of sodium sulphate. The filtrate was sealed and kept in dry place. Oil and hexane mixture was later separated into oil and hexane, using rotary evaporator by distilling the hexane vapors under vacuum. The extracted oil is then poured in vials, labeled and stored for FAC and carotene analysis.

For the fatty acid methyl determination, 0.05g of the extracted oil was measured and 1.9ml of solvent hexane added and homogenized. Sodium methoxide of 0.1ml was also added until a cloudy solution is formed. After 10 minutes, the clear portion of the methyl ester was pipetted into gas chromatograph vials and covered with Teflon vial caps for analysis. The separation of the FAME by the GC machine was under programmed conditions. The fatty acids components were represented by the peaks on the chromatograph.

3.3 Statistical Analyses

3.3.1 Variability Profile

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 traits were calculated using SPSS statistical tool. This was done in order to know the extent of variation in the germplasm accessions. The average of the quantitative data was also standardized to give equal

UNIVERSITI SAINS ISLAM MALAYSIA
ISLAMIC SCIENCE UNIVERSITY OF MALAYSIA

weight to all measurements using the following formula (Microsoft Office Excel, 2007):

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

Where, X = Value to standardize

μ = Arithmetic mean

σ = standard deviation of the distribution

3.3.2 Principal Components Analysis (PCA)

PCA simplifies the complex data by transforming 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 (CAMO software version 10.1). Mathematically, PCA involved in the decomposition of 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 data is the combination of scores and loadings in which focused by user in interpretation of PCA results while the remaining part is called error or residual matrix. The ath column of T and ath row of P is represented by vectors of t_a and p_a respectively and both are the vector representations of the ath 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 (CAMO, 2014).

In contrast, the residual matrix, *E*, represent the fraction of variation that cannot be modeled well. It other words, it cannot be explained by available PCs, yet, useful in the lack-of-fit measurement of model to the original data where small value of *E* indicated the good model and vice-versa (CAMO, 2014). Usually after PCA, the size of each component can be measured and represented by eigenvalue. In PCA, the more significant the components indicated the larger of their size, thus have larger eigenvalue. Therefore, the eigenvalue of a PC is the sum of squares of the scores and represented as follows:

$$g_a = \sum_{i=1}^{I} tij^2 \tag{3}$$

where g_a is the ath eigenvalue. The sum of all nonzero eigenvalues for a data matrix equals the sum of squares of the entire data-matrix, so that

$$\sum_{a=1}^{k} g_a = \sum_{i=1}^{I} \sum_{j=1}^{J} x_{ij}^2 \tag{4}$$

where *K* is the smaller of *I* or *J* (Brereton, 2003).

3.3.3 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 (CAMO software version 10.1). The general purpose of cluster analysis is to group similar objects or samples into respective classes based on their specified

characteristics or variables and it includes different type of algorithms such as joining (tree clustering), two-way joining (block clustering) and *k*-means clustering. The aim of the tree clustering algorithm was to join the objects or samples in each class of themselves together using specified distance measures and linkage rules, thus, forming larger cluster by connecting all the objects at the last step known as hierarchical tree.

The Ward's method (Ward, 1963) 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:

$$SSE = \sum_{i=1}^{k}\sum_{j=1}^{n_i}(y_{ij} - \bar{y})^2 \qquad (5)$$

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.

UNIVERSITI SAINS ISLAM MALAYSIA
ISLAMIC SCIENCE UNIVERSITY OF MALAYSIA

CHAPTER IV

RESULTS

4.1 Variability Profile

The oil palm germplasm from Senegal and Gambia exhibited low coefficient of variation to high coefficient of variation for the various plant attributes investigated (Table 1). A high variance for fresh fruit bunch (FFB), carotene, leaf number (LN), mesocarp-to-fruit ratio (M/F), shell-to-fruit ratio (S/F) and oil yield (OY) was observed, whereas for the rest of the other traits a low to medium variation was observed.

Table 1: Extent of Variation.

Characters	Coefficient of variation (%)
Yield traits	24.36 - 197.68
Bunch quality traits	0.69 - 85.06
Vegetative traits	0.50 – 107.20
Physiological traits	0.76 - 34.30
Fatty acid composition	0 – 59.09

4.2 Principal Components Analysis (PCA)

In order to determine with which combination type of the various agronomic traits the Senegal and Gambian oil palm germplasm would achieve high yield, PCA was performed. The result of the PCA therein explained the genetic diversity of the oil palm germplasm.

UNIVERSITI SAINS ISLAM MALAYSIA
ISLAMIC SCIENCE UNIVERSITY OF MALAYSIA

Table 2: Variability profile of Senegal and Gambian oil palm germplasm as analyzed by descriptive statistics

Parameter	N	Minimum	Maximum	Mean	Std. Deviation	Variance
FFB	44	24.36	197.68	84.59	33.40	1115.64
BNO	44	8.82	25.16	18.45	3.40	11.58
ABW	44	2.75	13.26	4.62	2.06	4.23
BWT	44	0.90	13.81	4.12	2.24	5.03
MFW	44	1.93	12.50	3.03	1.95	3.80
MNW	44	1.24	2.50	1.61	0.23	0.05
PB	44	0.00	5.10	0.46	0.81	0.66
MF	44	28.09	85.06	38.02	11.20	125.36
KF	44	6.06	19.15	15.00	2.75	7.56
SF	44	8.88	58.62	46.98	9.27	85.87
ODM	44	61.73	77.77	70.18	3.55	12.58
OWM	44	32.13	49.68	41.16	4.00	15.99
FB	44	47.63	65.63	57.74	3.35	11.23
OB	44	4.58	23.33	9.31	3.57	12.76
KB	44	3.49	11.53	8.65	1.70	2.90
OY	44	0.69	45.76	8.82	8.87	78.71
KY	44	0.94	10.45	6.83	2.13	4.52
TEP	44	1.25	50.05	12.92	9.20	84.62
FP	44	24.00	34.40	31.03	2.34	5.45
PCS	44	8.98	32.85	17.27	4.51	20.36
RL	44	3.20	5.33	3.93	0.38	0.15
LL	44	74.55	107.20	86.27	5.86	34.31
LW	44	3.46	5.16	4.05	0.39	0.15
LN	44	105.00	169.00	125.81	11.68	136.52
HT	44	2.14	3.50	2.81	0.35	0.12
LA	44	3.84	8.49	5.05	0.97	0.94
LAI	44	2.27	5.03	2.99	0.57	0.33
DIAM	44	0.50	0.86	0.68	0.06	0.00
F	44	0.60	0.89	0.70	0.06	0.00
LAR	44	10.09	18.56	13.01	1.65	2.73
BDM	44	0.67	15.84	6.59	2.76	7.63
VDM	44	7.27	19.52	12.26	2.35	5.51
TDM	44	12.26	34.30	18.85	4.17	17.40
BI	44	0.05	0.49	0.34	0.09	0.01
E	44	0.56	1.25	0.87	0.13	0.02
NAR	44	8.13	16.18	12.37	1.63	2.65
C14:0	42	0.21	1.13	0.50	0.15	0.02
C16:0	42	32.21	44.76	39.22	2.33	5.43
C16:1	42	0.07	0.58	0.15	0.10	0.01
C18:0	42	2.20	7.25	5.15	0.81	0.65
C18:1	42	38.63	53.93	44.77	3.05	9.28
C18:2	42	6.43	13.29	9.88	1.25	1.56
C18:3	42	0.10	0.60	0.19	0.09	0.01
C20:0	42	0.00	0.29	0.12	0.07	0.00
IV	42	53.27	59.09	56.26	1.50	2.26
Carotene	42	391.67	2405.88	1650.86	398.73	158987.16

UNIVERSITI SAINS ISLAM MALAYSIA
ISLAMIC SCIENCE UNIVERSITY OF MALAYSIA

4.2.1 Scores Plot

The score plot explained the percentage variance associated with each principal component obtained by drawing a graph between eigenvalues and principal component numbers. Nine principal components (PCs) exhibited eigenvalue of more than one and accounted for 88% of variability and as a result, these nine PCs were given due importance for further explanation (Figure 3). PC 1 showed 28% of variability with eigenvalue 11.20 in germplasm which then decreased gradually. Elbow line is obtained after which after 4th PC tended to straight. After that, little variance could be observed in each PC and it ended at 9th PC with variability of 3% and eigenvalue of 1.21.

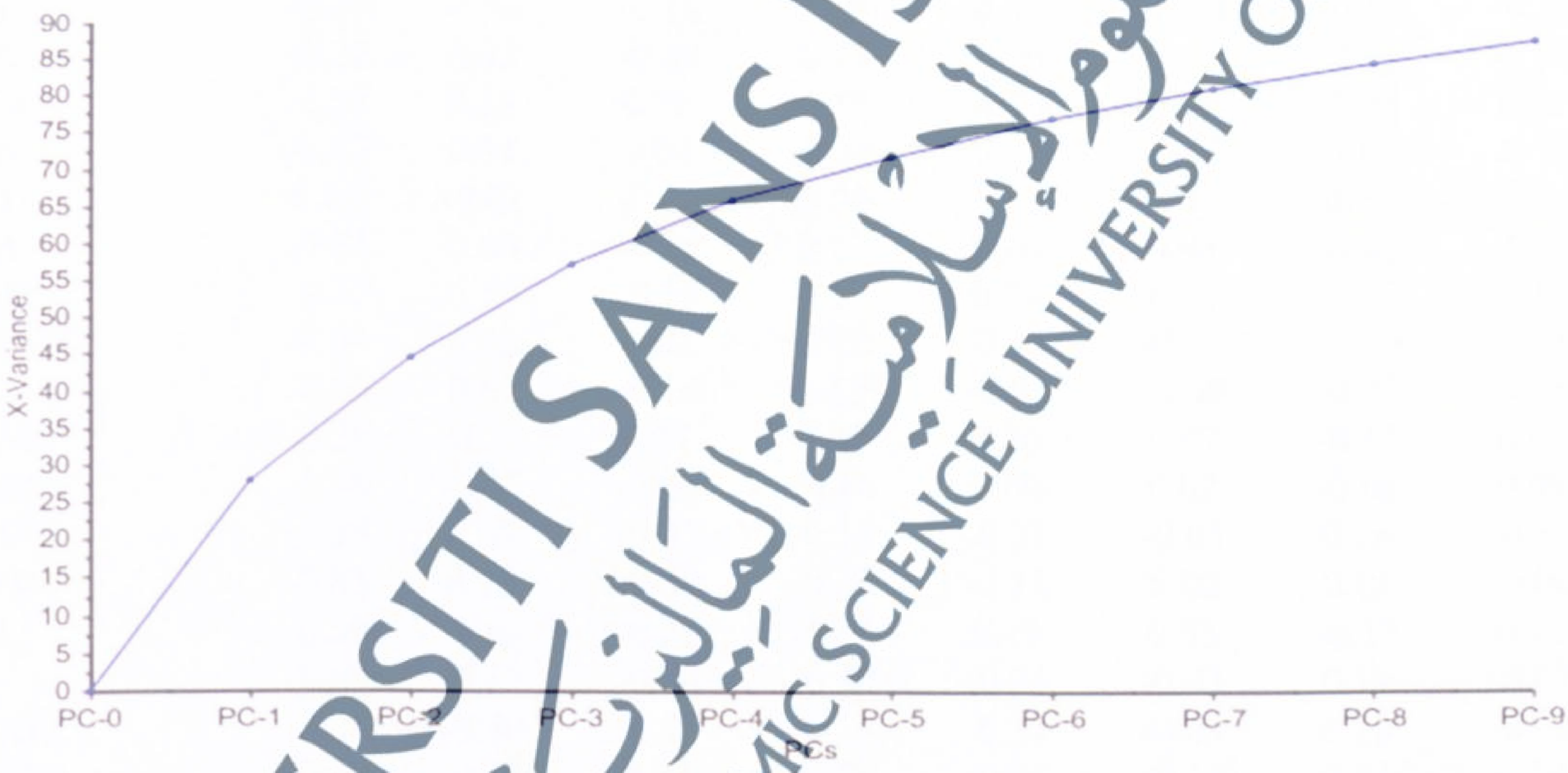


Figure 3: Score plot of principal component analysis between percentage variance and number of principal components

4.2.2 Loadings

The result of the PCA shown in Table 2 indicates the contribution of the various agronomic traits across the extracted PCs.

UNIVERSITI SAINS ISLAM MALAYSIA
جامعة العلوم الإسلامية الماليزية
ISLAMIC SCIENCE UNIVERSITY OF MALAYSIA
USIM

Table 3: Principal component analysis for Senegal and Gambia oil palm germplasm based on 46 agronomic traits

Traits	Principal Component Axes								
	1	2	3	4	5	6	7	8	9
FFB	-0.89	-0.29	-0.22	-0.09	-0.09	0.04	-0.13	0.13	-0.09
BNO	-0.22	-0.60	-0.65	-0.15	-0.08	-0.05	-0.10	0.14	-0.10
ABW	-0.96	-0.07	0.13	0.00	-0.05	0.10	-0.10	0.08	-0.01
BWT	-0.96	-0.10	0.14	-0.10	0.07	0.05	-0.02	0.03	-0.03
MFW	-0.88	0.03	0.35	0.06	-0.04	0.27	-0.04	-0.04	-0.06
MNW	-0.21	-0.14	-0.01	0.23	-0.26	0.74	-0.02	0.20	0.04
P/B	-0.83	0.06	0.24	-0.15	0.01	0.11	-0.18	-0.19	-0.15
M/F	-0.84	-0.04	0.47	0.03	0.07	0.13	-0.03	-0.15	-0.03
K/F	0.66	-0.06	-0.25	-0.22	-0.14	0.08	-0.32	-0.07	0.49
S/F	0.78	0.07	-0.48	0.03	-0.04	-0.18	0.15	0.20	-0.13
O/DM	-0.39	-0.18	0.51	-0.10	-0.01	-0.55	0.14	-0.09	0.09
O/WM	-0.39	-0.20	0.51	-0.09	0.05	-0.53	0.08	-0.21	0.11
F/B	0.24	-0.27	0.44	0.13	-0.34	0.49	0.01	-0.16	0.19
O/B	-0.75	-0.16	0.56	0.04	0.05	0.12	0.00	-0.20	0.07
K/B	0.65	-0.16	-0.07	-0.13	-0.25	0.26	-0.28	-0.12	0.52
OY	-0.94	-0.14	0.25	-0.05	0.01	0.11	-0.05	-0.04	-0.02
KY	-0.30	-0.61	-0.53	-0.31	-0.07	0.14	-0.17	0.06	0.25
TEP	-0.94	-0.24	0.15	-0.10	0.00	0.13	-0.07	-0.03	0.03
FP	-0.09	-0.64	-0.27	-0.15	0.00	-0.12	0.39	-0.21	0.22
PCS	-0.45	0.78	-0.26	-0.11	-0.11	-0.03	0.07	-0.06	0.12
RL	-0.49	0.74	-0.10	-0.20	0.07	-0.10	-0.11	-0.06	0.05
LL	-0.26	0.43	-0.50	0.17	0.36	-0.08	-0.14	-0.06	0.17
LW	-0.58	0.33	0.12	0.38	-0.38	0.08	-0.08	0.24	-0.13
LN	-0.47	0.51	0.03	-0.26	0.13	-0.06	0.14	-0.20	0.29
HT	0.08	-0.02	-0.26	0.06	-0.31	0.32	0.65	-0.14	-0.22
LA	-0.74	0.60	-0.15	0.17	0.04	0.03	-0.02	-0.01	0.12
LAI	-0.75	0.59	-0.15	0.17	0.04	0.03	-0.02	-0.01	0.12
DIAM	-0.04	0.72	0.07	0.02	-0.52	-0.11	-0.10	-0.08	0.06
F	-0.63	0.67	-0.16	0.15	-0.02	-0.06	-0.03	0.03	0.17
LAR	-0.39	-0.57	0.11	0.31	0.50	0.07	-0.15	0.07	0.19
BDM	-0.88	-0.32	-0.26	-0.16	0.00	0.02	-0.05	0.09	-0.06
VDM	-0.38	0.69	-0.35	-0.12	-0.31	-0.03	0.26	-0.17	0.16
TDM	-0.85	0.18	-0.39	-0.19	-0.18	0.00	0.12	-0.04	0.05
BI	-0.59	-0.68	-0.26	-0.16	0.19	0.03	-0.12	0.15	-0.03
E	-0.69	-0.17	-0.46	-0.40	-0.26	-0.03	0.19	-0.07	-0.04
NAR	-0.25	-0.50	-0.43	-0.55	-0.31	-0.09	0.20	-0.08	-0.11
C14:0	-0.29	-0.49	0.42	0.39	-0.01	-0.17	0.22	0.30	0.06
C16:0	-0.17	-0.30	-0.41	0.76	-0.10	-0.09	0.02	-0.28	0.03
C16:1	-0.09	0.17	-0.65	0.45	0.30	-0.09	-0.04	0.02	-0.12
C18:0	0.30	0.00	0.41	-0.29	-0.30	-0.03	-0.44	-0.25	-0.35
C18:1	0.16	0.43	0.11	-0.73	0.33	0.20	0.05	0.24	-0.04
C18:2	-0.22	-0.44	0.27	0.49	-0.41	-0.28	0.12	0.04	0.25
C18:3	-0.12	0.19	-0.63	0.31	-0.02	-0.06	-0.04	0.22	0.02
C20:0	-0.05	-0.34	-0.29	-0.24	-0.30	-0.33	-0.44	-0.10	-0.04
IV	-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

UNIVERSITI SAINS ISLAM MALAYSIA
ISLAMIC SCIENCE UNIVERSITY OF MALAYSIA

The most contributing agronomic traits to PC 1 with their corresponding eigenvector in bracket were: FFB (-0.89), ABW (-0.96), BWT (-0.96), MFW (-0.88), P/B (-0.53), M/F (-0.84), S/F (0.78), K/F (0.66), O/B (-0.75), K/B (0.65), OY (0.66), TEP (-0.94), BDM (-0.89), TDM (-0.85), and *e* (-0.69). On PC 2, the most contributing traits with their corresponding eigenvector in bracket were BNO (-0.60), KY (-0.61), FP (-0.64), PCS (0.78), RL (0.74), LA (0.60), DIAM (0.72).

For PC 3 BNO (-0.65), O/DM (-0.51), O/WM (0.57), O/B (0.56), KY (-0.53), LL (-0.50), C16:1 (-0.55), and C18:3 (-0.63) were the most contributing traits. The most contributing traits to PC 4, are PC 5, PC 6, PC 7, PC 8 and PC 9 with their corresponding eigenvectors in bracket were: NAR (-0.55), C16:0 (0.76), C18:1 (-0.73), IV (-0.50): DIAM (-0.52), carotene (0.56): O/DM (-0.55), O/WM (-0.53): HT (0.65): IV (0.52), carotene (-0.51): and K/B (0.52) respectively.

Figure 4 is a PC loading plot and has two dimensional scatter plots showing the distribution of the various traits in space. It could be observed from the graph that each variable has a loading on the plot. Variables along the vertical axis are the ones with the highest contribution to the PC 1 while those along the horizontal axis are correlated to the PC 2. It is also visible on the graph that some of the variables are superimposed on one another while some of the variables on the other hand are further apart on two opposite ends.

The correlation loadings plot of the various traits in two dimensional axes is shown in Figure 5. The variables on the inner circle indicate 50% variance while those found on the outer circle indicate 100% variance and they exhibit high contribution to PC 1.

UNIVERSITI SAINS ISLAM MALAYSIA
ISLAMIC SCIENCE UNIVERSITY OF MALAYSIA

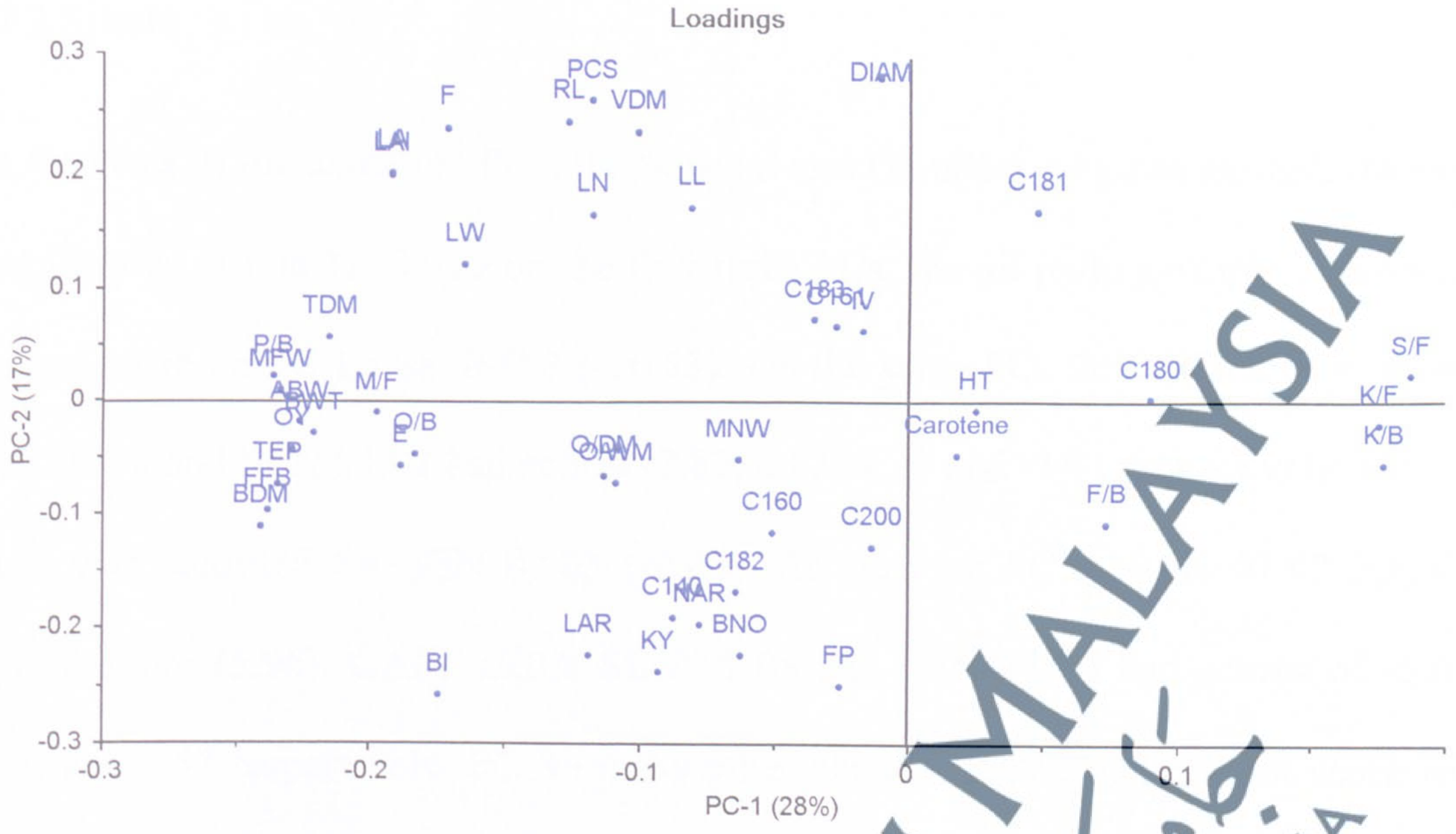


Figure 4: Scattered diagram of 46 oil palm germplasm traits for first two components contributing almost half of the total variability

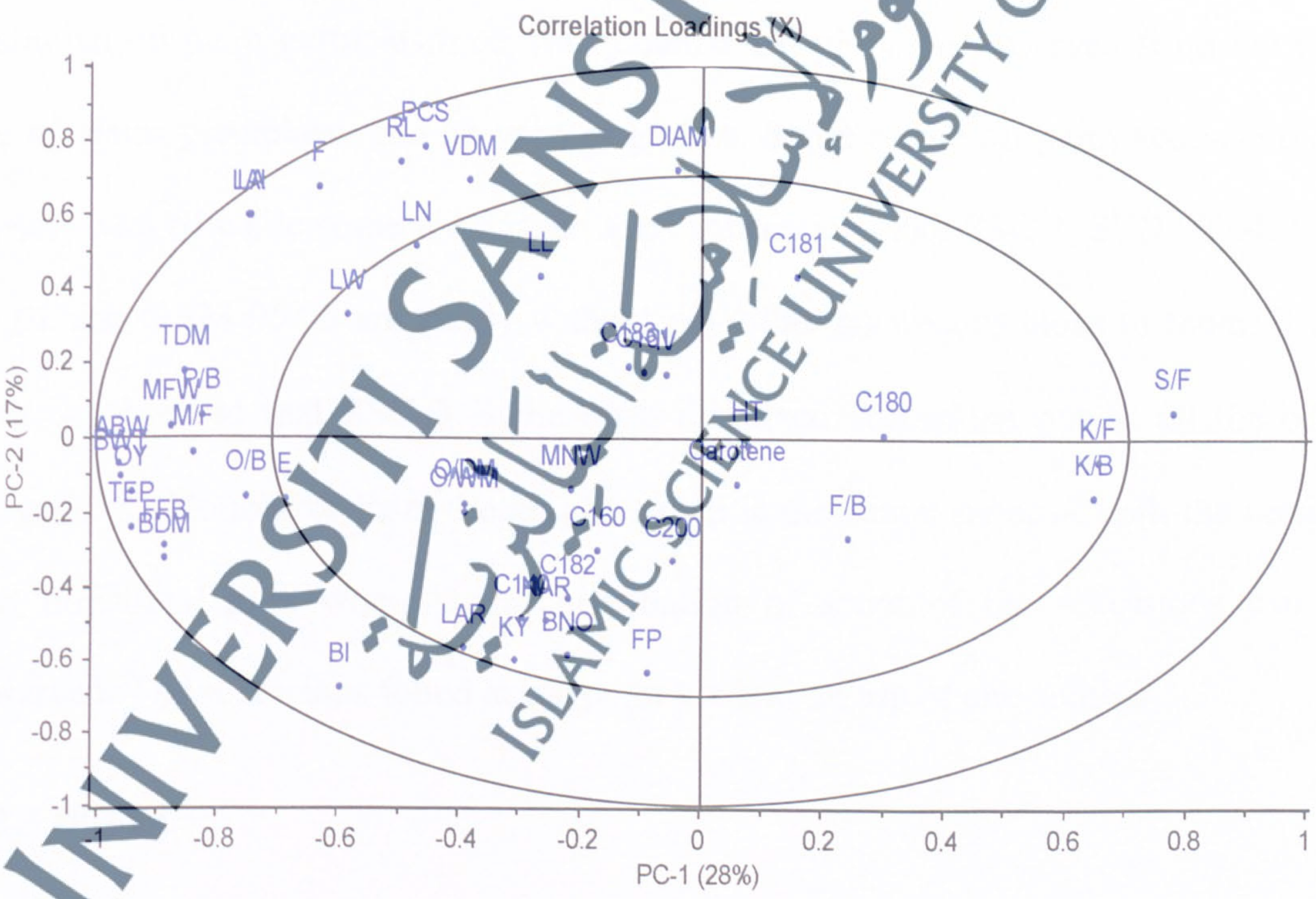


Figure 5: Scattered diagram of 46 oil palm germplasm traits showing correlation to the first two components

UNIVERSITI SAINS ISLAM MALAYSIA
ISLAMIC SCIENCE UNIVERSITY OF MALAYSIA

4.2.3 Scores

On the basis of the extracted PCs, the Senegal and Gambian oil palm germplasm were given scores (Table 3). Based on the first three PCs, the oil palm germplasm with the highest score on PC 1 was SSC 3 (-16.53). On the same PC, SEN 09.04, SEN 13.07, SEN 02.08 and SEN 01.02 had scores -7.87, 4.83, 4.25 and -3.11 respectively. Lowest score was recorded for SEN 07.03 (-0.03). As seen on PC 2, SEN 01.02 had the highest score (5.96). GAM 05.08, SEN 02.08 and SEN 13.04 had scores of -5.64, 5.04, and 4.57 respectively. PC 3 on the other hand, revealed that highest score was recorded by SEN 01.02 (-8.83) while the least was by SEN 10.05 (-0.02) and SEN 05.02 (-0.02).

The scatter plot as displayed in Figure 6 shows the distribution of the Senegal and Gambian oil palm germplasm on two dimensional plots. As observed from the plot, the oil palm germplasm can be seen in groups. Some of the oil palm accessions are closely packed while some are further apart. Accessions like SSC 3, SEN 09.04, SEN 01.02 and GAM 05.08 are singly without any other accessions close to them. It can also be observed that SSC 3 is the most dispersed accessions out of all the other accessions. Noticeable from the graph as well is the center point of both the vertical and horizontal axes where dense distribution of some of the accessions can be observed. The accessions found at the point are laid on top of one another.

4.2.4 Bi-plot

A PC bi-plot shows the simultaneous distribution of the traits as well as the accessions with the high scores for the traits (Figure 7). On the graph, it can be seen that the

Senegal and Gambian the oil palm accessions are positioned close to the traits that best portray them.

Table 4: 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

UNIVERSITI SAINS ISLAM MALAYSIA
ISLAMIC SCIENCE UNIVERSITY OF MALAYSIA

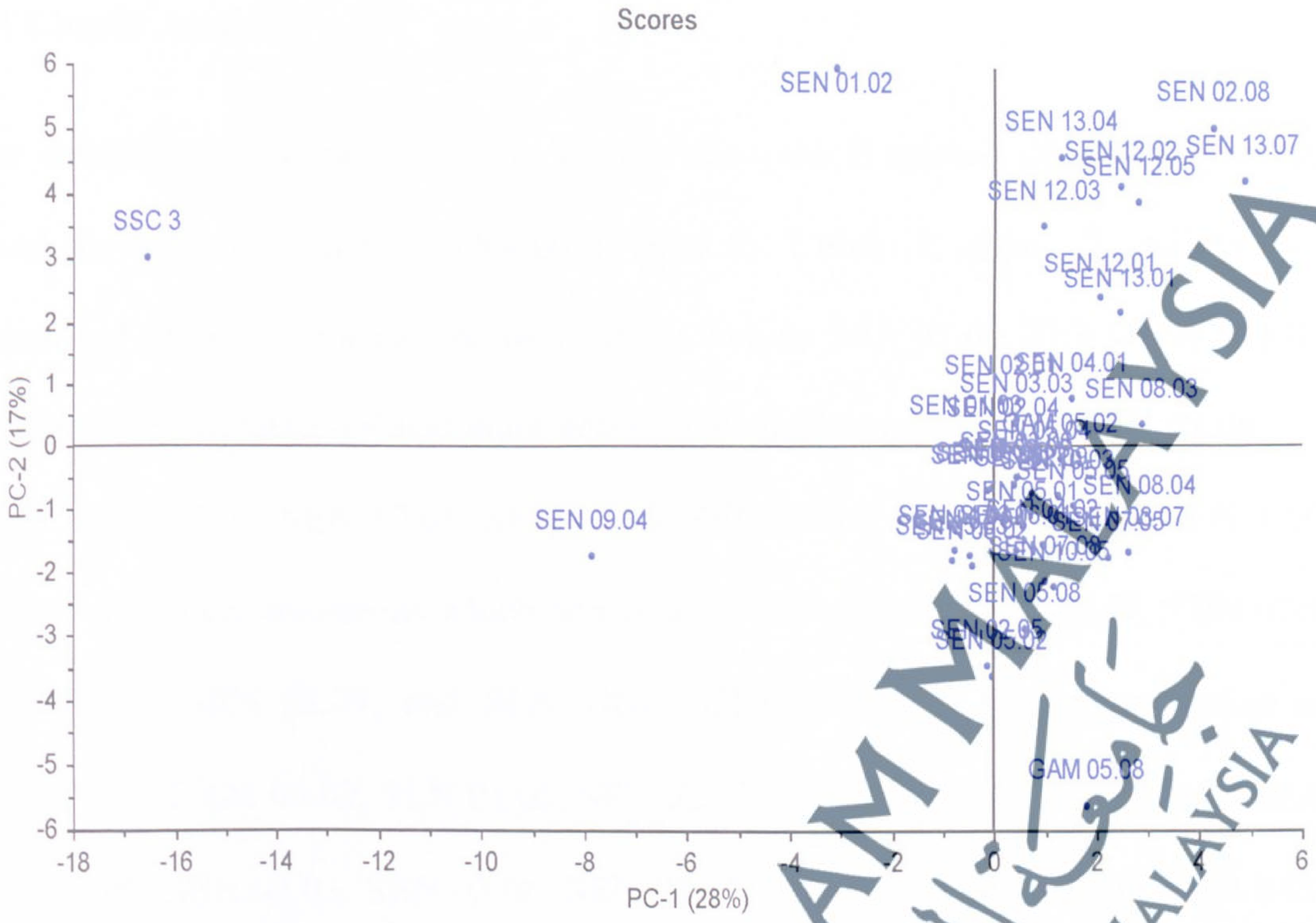


Figure 6: Two dimensional ordinations of 42 Senegal-Gambian oil palm accessions on principal axes 1 and 2

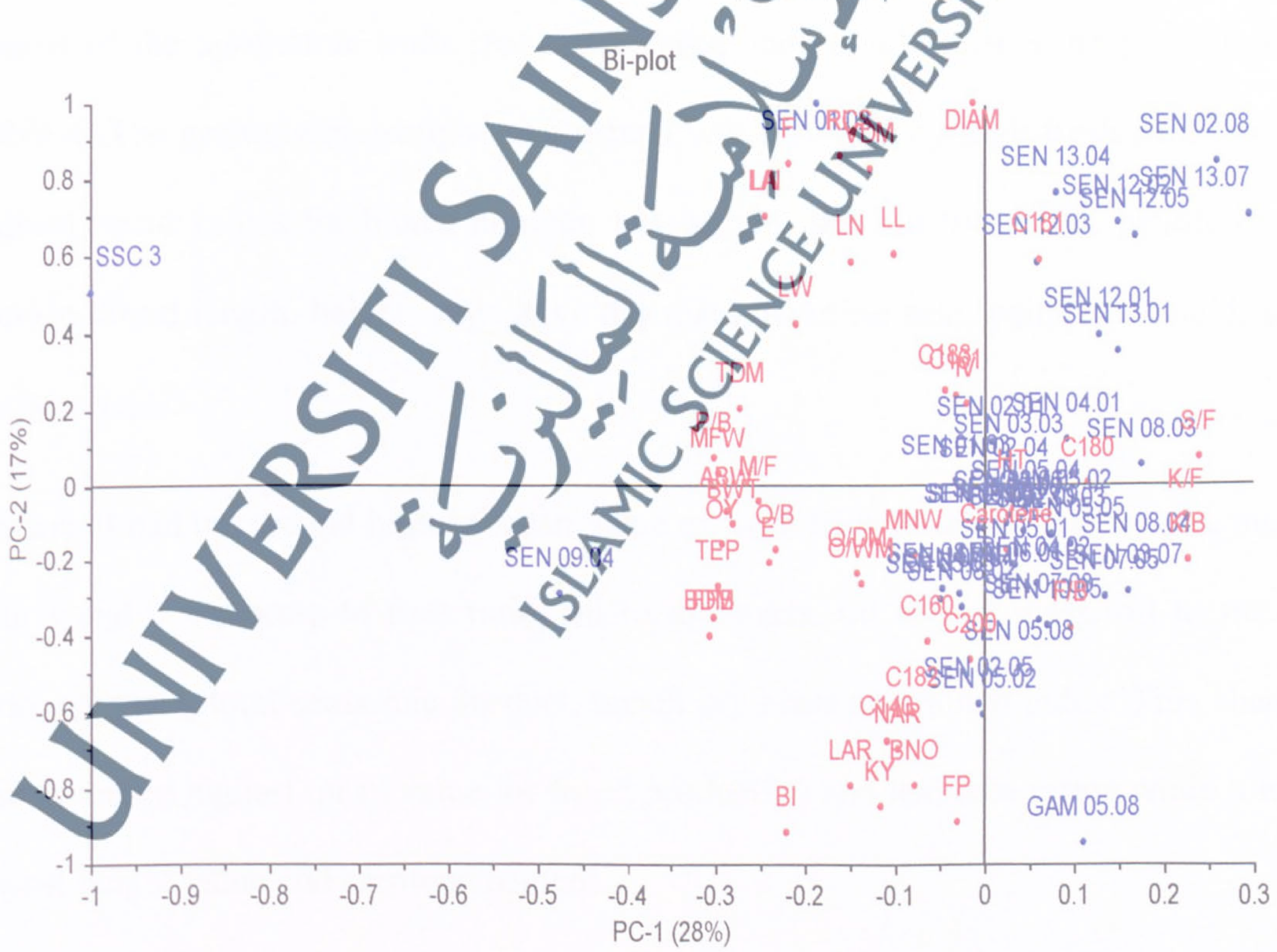


Figure 7: Bi-plot of 46 oil palm agronomic traits and 42 oil palm accessions on PC 1 and PC 2

UNIVERSITI SAINS ISLAM MALAYSIA
ISLAMIC SCIENCE UNIVERSITY OF MALAYSIA
جامعة العلوم الإسلامية الماليزية

4.3 Cluster Analysis

The dendrogram generated through Ward's hierarchical method divided the Senegal-Gambian accessions into six clusters (Figure 8). Cluster-I, cluster-II and cluster-III comprised of only one accession each which include SEN 01.02, SEN 09.04 and SSC 3 respectively. Cluster-IV had eight accessions which comprised of SEN 02.08, SEN 12.01, SEN 12.02, SEN 12.03, SEN 12.05, SEN 13.01, SEN 13.04, and SEN 13.07. Cluster-V had six accessions which consisted of GAM 05.08, SEN 05.02, SEN 07.05, SEN 08.03, SEN 08.04, and SEN 10.03. Cluster-VI was the largest group and consisted of GAM 05.02, SEN 01.03, SEN 02.01, SEN 02.04, SEN 02.05, SEN 02.06, SEN 02.09, SEN 03.03, SEN 03.06, SEN 03.07, SEN 04.01, SEN 04.02, SEN 04.03, SEN 05.01, SEN 05.03, SEN 05.04, SEN 05.05, SEN 05.08, SEN 06.01, SEN 06.08, SEN 07.03, SEN 07.04, SEN 07.08, SEN 08.02, and SEN 10.05.

Means of the agronomic traits that characterize individual clusters are presented in Table 4. The germplasm grouped in cluster-I was reflected by high fresh fruit bunch, highest mean values for bunch number, nut weight, shell to fruit ratio, petiole cross section, frond length, height, vegetative dry mass, palmitic acid, palmitoleic acid, and linoleic acid.

Cluster-II had the second highest mean value of fresh fruit bunch, bunch weight, mean fruit weight, mesocarp-to-fruit ratio, oil to dry mass, oil to wet mass, oil to bunch ratio, oil yield, total economic product, bunch dry mass and bunch index. This cluster also recorded highest mean value for frond production and leaf area ration while it had lowest height value and carotene content.

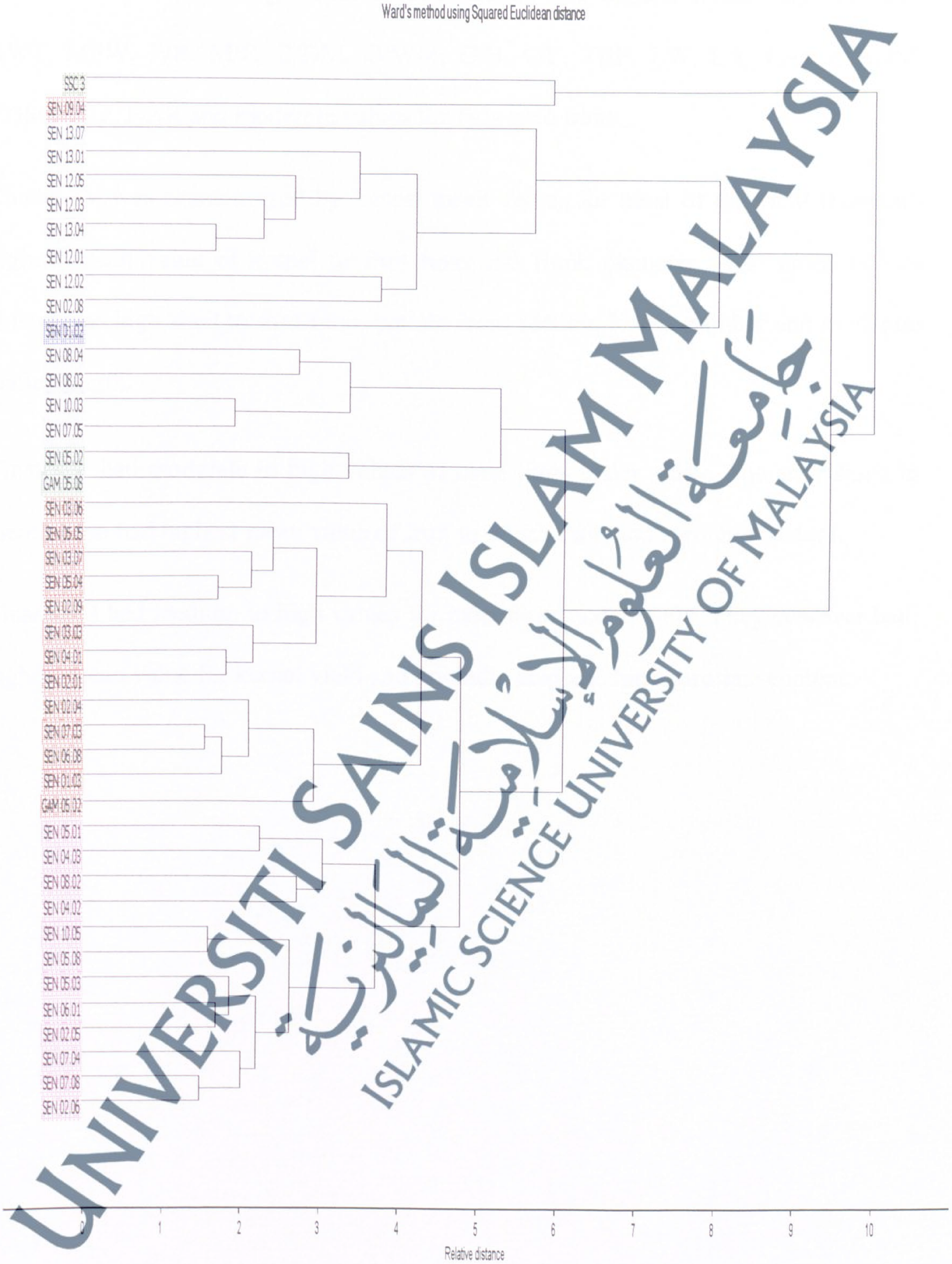


Figure 8: The relationship among the oil palm germplasm accessions reflected by cluster analysis

Cluster-III was noted for having the highest mean value for most of the yield traits as well as morpho-physiological traits. This cluster had highest mean value of FFB, BWT, MFW, P/B, M/F, O/DM, O/WM, O/B, OY, TEP, LW, LA, LAI, *f*, BDM, TDM, BI, *e*, NAR and moderate values for fatty acid traits.

Cluster-IV was characterized by lowest mean values for most of the yield traits and highest mean value of kernel to fruit ratio and trunk diameter. Also noted in this cluster was high shell to fruit ratio, petiole cross section, leaflet number and moderate leaflet length.

Cluster-V had moderate to high values of most traits under study. The accessions in therein also had highest mean value of fruit to bunch ratio and carotene content.

Cluster-VI had medium to high values for most traits under study. They however had, highest mean value for kernel yield and arichidic acid and high carotene content.

UNIVERSITI SAINS ISLAM MALAYSIA
ISLAMIC SCIENCE UNIVERSITY OF MALAYSIA

Table 5: 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

UNIVERSITI SAINS ISLAM MALAYSIA
ISLAMIC SCIENCE UNIVERSITY OF MALAYSIA
YSIA

4.3.1 Genetic Distance

The genetic distance of all the oil palm germplasm as analyzed by the proximity matrix of squared Euclidean distance (see Appendix 1), revealed that the largest genetic distance was between SEN 13.07 and SSC 3 (520.302) followed by SEN 02.08 and SSC 3 (454.034). The least distance was found between SEN 02.06 and SEN 07.08 (10.366). The inter cluster distance as shown in Table 5 also revealed that the largest genetic distance was between cluster-III and cluster-V (158.629) while the least was between cluster-V and cluster-VI.

Table 6: 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

UNIVERSITI SAINS ISLAM MALAYSIA
ISLAMIC SCIENCE UNIVERSITY OF MALAYSIA
USIM

CHAPTER V

DISCUSSION

A clear relationship between agronomic traits as stated by Deyong (2011) will not only address the fundamental principle of plant breeding but will also facilitate the plant breeding practice. Most of the characters such as the agronomic traits investigated in the current study contribute to crop yield. The high variability unveiled in some of the yield, morpho-physiological and fatty acid traits are essential and important for efficient selection of elite oil palm accessions with high yield and yield component characters. This finding supports the result reported by Li-Hammed et al. (2014) in variation and correlation studies of MPOB-Nigerian oil palm germplasm. The existence of variation is nonetheless vital for diversity and adaptability in breeding population (Marhalil et al., 2013).

5.1 Principal Component Analysis (PCA)

Deyong (2011) 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 conception according to Deyong (2011) could be materialized through multivariate statistical analysis. 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 ABW and BWT, PCS, BNO, C16:0, carotene, MNW, HT, IV and K/B

contributing mostly to PC1, PC 2, PC 3, PC 4, PC 5, PC 6, PC 7, PC 8 and PC 9 respectively. This translates that oil palm accession with higher scores for these traits seems probably will attain high yield more easily and as a result, these traits should be given utmost importance in breeding program of oil palm germplasm under study. This result follows similar trends to that of Deyong (2011) 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 was poor on other PCs except on 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 of Maqbool et al. (2010) on morphological diversity and traits association in bread wheat.

5.1.1 Score Plot

The score plot 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 (Fernadez, 2002). The score plot in this study showed that maximum variation was present in the first PC and according to Maqbool et al. (2010), selection of genotypes from PC 1 will be useful.

5.1.2 Loadings

PC loadings are correlation coefficients between the PC scores and the original variables. PC loadings measure the importance of each variable in accounting for the

UNIVERSITI SAINS ISLAM MALAYSIA
ISLAMIC SCIENCE UNIVERSITY OF MALAYSIA

variability in the PC (Fernandez, 2010). 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 be anti-correlated. That is to say, 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 palm 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 Li-Hammed et al. (2014) on association between oil palm traits.

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 (CAMO, 2014). As stated by Fernandez (2010), 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 dataset al. so, 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 maybe of little importance (Fernandez, 2010).

5.1.3 Scores

PC scores are the derived composite scores computed for each observation based on the eigenvectors for each PC (Fernandez, 2010). From the scores of the oil palm germplasm summarized in Table 3, 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 (CAMO, 2014). 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 (38 traits) while the non-yield traits were also positively loaded on PC 1(for 11 traits) 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 programs.

Furthermore, score 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, oil palm genotypes that are closer to one another have close values for the corresponding variables while those that are far away from one another are quite different in values for corresponding variables (CAMO, 2014). 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 (Smiullah, 2013).

In the present study, SSC 3 was the most distant oil palm genotype from other oil palm accessions. The reason for this is glaring distance 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 (*pisifera* is normally trail) used in this study (Teoh, 2002; Corley & Tinker, 2003). Besides 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.

5.1.4 Bi-Plot

Bi-plot display is a visualization technique for investigating the inter-relationships between the observations and variables in multivariate data (Fernendez, 2010). From the bi-plot (Figure 5), genotypes SSC 3, SEN 09.04 and SEN 01.02 will be 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. (2013) who used the bi-plot graph to suggest good candidates of cowpea accessions to be used in genetic improvement of the crop.

5.2 Cluster Analysis

Estimating genetic diversity and determining the relationships between collections are very useful for ensuring for ensuring efficient germplasm collections and different markers are available for studying variability among accessions (Rabbani et al., 1998). 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 (Mostafa & Hamed, 2011).

UNIVERSITI SAINS ISLAM MALAYSIA
ISLAMIC SCIENCE UNIVERSITY OF MALAYSIA

Doumbia et al. (2013) stated that 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. Cluster analysis of the germplasm resources is helpful for parental selection in the plant breeding program (Deyong, 2011). In this study, all the Senegal and Gambia oil palm germplasm were clustered into six types with each group having its own peculiar characters. 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 (Li-Hammed et al. 2014).

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. Seymus and Brulent (2010), Lacis et al. (2010), Talebi and Rockzadi (2013) and Ajmal et al. (2013) also reported lack of relationship between geographical origin and distance. Sonnante and Pignone (2007) were of the opinion that the association between genetic similarity and geographic distance among genotypes is not always clear. This may be due to migration of the oil palm materials from one region to another in collection site.

Based on the means value for each cluster, cluster-I, cluster-II and cluster-III contained genotypes with high yield characters and as stated by Ajmal et al. (2013), increased yield potential is a stated goal for plant breeders. 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. (2013). 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 (Sapey et al., 2012). Groups with desired traits can also be exploited directly for such traits.

5.2.1 Genetic Distance

According to squared Euclidean distances (D^2) among genotypes, 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. WeldeMicheal et al. (2013) reported maximum heterosis in some germplasm accessions of some Ethiopian specialty coffee based on morphological traits.

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 close proximity should be avoided. However, Rahim et al. (2010) suggested that crosses between close genotypes could be useful for backcross breeding programs. Similar to the findings of Khodadadi et al. (2010), who reported that cluster analysis can be used for finding high yielding genotypes in wheat.

CHAPTER VI

CONCLUSIONS AND RECOMMENDATIONS

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. The characters that were dominants in the first components are closely related to yield and yield components; while vegetative and physiological traits like the radiation conversion efficiency (*e*), fractional interception of radiation (*f*) and leaf characters were associated with second components and other extracted components. The fatty acid traits were not that dominant on the extracted PCs. This recommends the likelihood of attaining, through selection, suitable genotypes combining high yield with desirable traits for direct release as cultivars in the Malaysian Palm Oil Board (MPOB).

Cluster analysis also aided in distinguishing accessions on the basis of their different levels of similarity. Six groups were identified with precise differences according to the extracted principal components. Compared with other groups, cluster-III showed highest mean values for yield traits. With the exception of cluster-IV which had lowest mean value for yield traits, all other groups had moderate to high values of yield and yield related traits. The cluster analysis however, provides valuable information in order to utilize directly the most promising accessions for production (for example SSC 3, SEN 09.04) or for future usage in selection programs. Furthermore, there was no consistency in assigning of members to groups, as accessions from different geographical origin were placed together.

UNIVERSITI SAINS ISLAM MALAYSIA
ISLAMIC SCIENCE UNIVERSITY OF MALAYSIA
USIA

However, caution must be taken as regard the accessions because they are an expression of linked genetic and environmental effects, i.e. findings were based on morphological traits which are usually influenced by environment. Therefore, further studies should be undertaken to ascertain the claims of this study. For example, molecular techniques which are more precise and are not affected by the environment can be conducted to further verify the findings from this study. The findings of this research will however be of utmost help to researchers, plant breeders and policy makers.

UNIVERSITI SAINS ISLAM MALAYSIA
ISLAMIC SCIENCE UNIVERSITY OF MALAYSIA

REFERENCES

Adebisi, M. A., F. S. Okelola, M. O. Ajala, T.O. Kehinde, I. O. Daniel & O.O. Ajani. 2013. "Evaluation of Variations In Seed Vigour Characters of WestAfrican Rice (*oryza sativa L.*) Genotypes Using Multivariate Technique". *American Journal of Plant Sciences*, Vol.4 (2), 2013. p. 356-363.

Ajmal, S., N.M. Minhas, A. Hamdani, A. Shakir, M. Zubair & M. Ahmad. 2013. "Multivariate Analysis of Genetic Divergence in Wheat (*Triticum aestivum*) Germplasm". *Pakistan Journal of Botany*. Vol. 45. (5): p. 1643-1648.

Ariyo, O.J. 1993. "Genetic Diversity in West African Okra (*Abelmoschus caillei* L. Chev.) Multivariate Analysis of Morphological and Agronomical Characteristics". *Genetic Res. Crop Evol*. Vol.40. p. 25-32.

Barcelos, E., P. Amblard, , J. Berthaud & M. Seguin. 2002. "Genetic Diversity and Relationship in American and African Oil Palm as Revealed by RFLP and AFLP Molecular Markers". *Pesq. Agropec. Bras., Brasília*. Vol. 37. (8): p. 1105-1114.

Blaak, G. 1967. *Oil Palm Prospection Four in the Bamenda Highlands of West Cameroon*. London: Internal report, Unilever.

Bozokalfa, M.K., D. Esiyok & K. Turhan. 2009. "Patterns of Phenotypic Variation in a Germplasm Collection of Pepper (*Capsicum annum* L.) from Turkey". *Spanish Journal of Agricultural Research*. Vol. 7. (1): p. 83-95.

Brereton, R.G. 2009. *Chemometrics for Pattern Recognition*. UK: John Wiley & Sons Ltd Publisher.

Bradley, P.S, O.L. Mangasarian & W.N. Street. 1998. "Clustering via Concave Minimization," in Advances in Neural Information Processing Systems, vol. 9, M.C. Mozer, M. I. Jordan, and T. Petsche, Eds. Cambridge, MA: MIT Press, 1997, pp. 368–374

Brereton, R.G. 2003. *Chemometrics: Data analysis for the laboratory and chemical plant*. England: John Wiley & Sons Ltd.

CAMO. 2011. *The Unscrambler® X 2009-2011*. Version 10.1 (32-bit). CAMO software AS.

CAMO. 2011. *What is Mmultivariate Analysis?* USA: CAMO Software AS.

CAMO. 2014. http://www.camo.com/resources/methods.html. (Assessed on 6th January 2014).

Chavelier, A. 1943. "Taxonomie, Biogeographie et Selection des Palmiers au *Genre Elaeis*". *Bot. Appl. Agric. Trop*. Vol. 23. (295).

Chong, C.L. 1994. "Chemical and Physical Properties of Palm Oil and Palm Kernel Oil" in *Selected Readings on Palm Oil And Its Uses*. Kuala Lumpur, Malaysia: Palm Oil Research Institute of Malaysia.

Corley, R.H.V. & P.B. Tinker. 2003. *The Oil Palm*. 4th Edition. UK: Wiley-Blackwell. ISBN: 978-0-632-05212-7.

Darvishzadeh R., M. Dalkani & A. Hassani. 2012. "Determination of the genetic Variation in Ajowan (*Carum Copticum L.*) Populations Using Multivariate Statistical Techniques". *Revista Ciencia Agronomica*, Vol. 43 (4): p. 698-705. December, 2013.

Deyong, Z. 2011. "Analysis among Main Agronomic Traits of Spring Wheat (*Triticum aestivum*) in Qinghai Tibet Plateau". *Bulgarian Journal of Agricultural Science*. Vol. 17. (5): p. 615-622.

Doumbia, I.Z., R. Akromali & J.Y. Asibuo. 2013. "Comparative Study of Cowpea Germplasms Diversity from Ghana and Mali using Morphological Characteristics". *Journal of Plant Breeding and Genetics*. Vol. 1. (3): p. 139-147.

Dransfield, J., N.W. Uhl, C.B. Asmussen, W.J. Baker, M.M. Harley & C.E. Lewis. 2005. "A New Phylogenic Classification of the Palm Family, *Arecaceae*". *Kew Bulletin*. Vol. 60. (40):p. 559-569.

Esbensen, K.H., D. Guyot, F. Westad & L.P. Houmoller. 2002. *Multivariate Data Analysis in Practice : an Introduction to Multivariate Data Analysis and Experimental Design* (Fifth Ed.). Norway: CAMO Process Publisher.

FAO. (1996) *Report on the state of the world's plant genetic resources for food and agriculture*. Rome: Food and Agriculture Organisation of the United Nations.

Fernandez G. 2010. *Principal component analysis*. RS701D. http://www.cabnr.unr.edu/saito/classes/ers701/pca2.pdf . (Assessed on 30th October 2014).

Franco, J., J. Crossa, M.L. Warburton, S. Taba & S.A. Eberhart. 2006. "Sampling strategies for conserving maize diversity when forming core subsets using genetic markers". *Crop Science*. Vol. 46. p. 854-864.

Gonzalez, A.G. 2012. "Critical Aspects of Supervised Pattern Recognition Methods for Interpreting Compositional Data" in *Chemometrics in Practical Applications*. Varmuza, K. (Ed.). China: Intech Publisher.

Hall R.I, P.R. Leavitt, R. Quinlan, A.S. Dixit & J.P. Smol. 1999. "Effects of Agriculture, Urbanization, and Climate on Water Quality in the Northern Great Plains". *Limnol. Oceanogr*. Vol. 44. (3-2): p. 739-756.

Hammer, K., N. Arrowsmith & T. Gladis. 2003. "Agrobiodiversity with emphasis on plant genetic resources". *Naturwissenschaften*. Vol. 90. p. 241-250.

Hamrick, J. L. & M.J.W. Godt. 1996. "Effects of life history traits on genetic diversity in plant species". *Philosophical Transactions of the Royal Society B: Biological Sciences*. Vol. 351. (1345): p. 1291-1298.

Hatcher, L. & E. Stepanski. 1994. *A step-by-step approach to using the SAS system for univariate and multivariate statistics*. Cary, NC: SAS Institute Inc.

UNIVERSITI SAINS ISLAM MALAYSIA
ISLAMIC SCIENCE UNIVERSITY OF MALAYSIA

Haydar, A., M.B. Ahmed, M.M. Hannan, M.A. Razvy, M.A. Manal, M. Salahin, R.Karim & M. Hossain. 2007. "Analysis of Genetic Diversily in Some Potato Varieties Grown in Bangladesh". *Middle East Journal of Scientific Research*. Vol. 2. (3-4): p. 143-145.

Hoon, M., S. Imoto & S. Miyano. 2013. *The Clustering Library for cDNA Microarray Data. Japan*. The University of Tokyo: Institute of Medical Science, Human Genome Center.

Hornovoka, O., M. Zavodna, M. Zakova, J. Kraic & F. Debre. 2003. "Diversity of Common Bean Landraces Collected in The Western And Eastern Carpatien". *Chech Journal of Genetic Plant Breeding*. Vol. 39. p. 73-83.

Jain, A. K. & R. C. Dubes, Algorithms for Clustering Data. Prentice-Hall, 1988.

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

Jalani, B.S. 2012. *Malaysian Oil Palm Industy, Contribution, Challenges and Future Prospects*. 71800, Bandar Baru, Nilai, Negeri Sembilan, Malaysia: Universiti Sains Islam Malaysia.

Jolliffe, I.T. 1986. *Principal Component Analysis*. Berlin: Springer-Verlag.

Jolliffe, I.T. 2002. *Principal Component Analysis*. (2nd Ed.). UK: Springer Publisher.

Karp, A., S. Kresovich, K. V. Bhat, W. G. Ayad & T. Hodgkin. 1997. *Molecular tools in plant genetic resources conservation: a guide to the technologies. IPGRI Technical Bulletin No. 2.* Rome, Italy: International Plant Genetic Resources Institute.

Khodadadi, M., M. H. Fotokian & M. Miransari. 2011. "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. (1): p. 17-24.

Lacis, G., V. Trajkovski & I. Rashid. 2010. "Phenotypical Variability and Genetic Diversity within Accessions of the Swedish Sour Cherry (*Prunus cerasus* L.) Genetic Resources Collection". *Biologija*. Vol. 56. (1-4): p. 1-8.

Latiff, A. 2000. "The biology of the Genus *Elaeis*" in *Advances in Oil Palm Research*. B. Yusof, B.S. Jalani, & K.W. Chan (ed.). Bangi: Malaysian Palm Oil Board, Bangi. p. 19-38.

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

Li-Hammed, M.A. 2014. *Principal Component Analysis and Clustering of Ex Situ Oil Palm* (*Elaeis guineensis* Jacq.) *Germplasm*. (MSc Thesis). Nilai, Universiti

Sains Islam Malaysia.

Lohani, M., D. Singh & J.P. Singh. 2012. "Genetic Diversity Assessment Through Principal Component Analysis in Potato (*Solanum tuberosum* L.)". *Vegetable Science*. Vol. 39. (2): p. 207-209.

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

Manpreet S., K. Keerat & S. Bhavdeep. 2008. "Cluster Algorithm For Genetic Diversity". *World Academy of Science, Engineering and Technology*. Vol. 18. (2008).

Maqbool, R, M. Sajjad & I. Khaliq. 2010. "Morphological diversity and traits association in bread wheat (*Triticum aestivum* L.)". *American-Eurasian Journal of Agriculture and Environmental Science*. Vol. 8. (2): p. 216-224.

Martens, H. & T. Naes. 1993. *Multivariate Calibration*. New York: Wiley.

Martinez-Calco, J., A.D. Gisbert, M.C. Alamar, R. Hernandorena, C. Romero, G. Ilacer & M.L Badenes. 2005. "Study of a germplasm collection of Loquat (*Eriobotrya japonica* Lindi) by multivariate alaysis". *Genetic Resources and Crop Evolution*. Vol. 55. (5): p. 697-703.

Matthias, O. 2007. Chemometrics. Weinheim: WILEY-VCH Verlag GmbH & Co.

Mostafa, A. & H. Felenji. 2011. "Evaluating Diversity among Potato Cultivars Using Agro-Morphological and Yield Components in Fall Cultivation of Jiroft Area". *American-Eurasian Journal of Agricultural and Environmental Science*. 11(5): 655-662.

Muhammad, A., S.A. Jatoi, T. Rafique & Abdul Ghafoor. 2013. "Genetic Divergence in Indigenous Spinach Genetic Resources for Agronomic Performance and Implication of Multivariate Analyses for Future Selection". *Science, Technology and Development*. Vol. 32. (2): p. 7-15.

Negri, V. & B. Tiranti. 2010. "Effectiveness of in situ and ex situ conservation of crop diversity. What a *Phaseolus vulgaris* L. landrace case study can tell us". *Genetica*. Vol. 138. p. 985-998.

Novembre, J. & S. Matthew. 2008. "Interpreting Principal Component Analysis of Spatial Population Genetic Variation". *Nature Genetics*. p. 646-649.

Oil world. 2008. Oil world 2007 annual report. Mielke, Hamburg.

Oil World.2012. *Oil world.* Hamburg, Germany.

Oyelola, B.A. 2004. (Paper). *The Nigerian Statistical Association Preconference Workshop*. Conference centre, University of Ibadan, Nigeria. 20-21 September.

Peerak S., T. Yimram & P. Somta. 2009. "Genetic Variation in Cultivated Mungbean germpalsm and its Implication in Breeding for High Yield". *Field Crops research*. Vol. 1. (12): p. 260-266.

Preston, J. 2011. *The characterization of heritable vegetable*. (PhD Thesis). University of Birmingham.

Purseglove, J.W. 1972. *Tropical Crops. Monocotyledons*. London: Longman Group Ltd., London. p. 479-510.

Rabbani M.A., A. Iwabuchi, Y. Murakami, T. Suzuki & K. Takayanagi. 1998. "Genetic Diversity of Mustard (*Brassica juncea* L.) Germplasm from Pakistan as Determined by RAPDs". *Euphytica*. Vol. 103. (2): p. 235-242.

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

Rajanaidu, N. & B.S. Jalani. 1994. "Oil Palm Genetic Resources, Collection, Evaluation, Utilization and Conservation". (Paper). PORIM *Colloquium on Oil Palm Genetic Resources*. PORIM, Bangi. 13 September.

Rajanaidu, N. 1994. "Oil palm cultivation and FFB production" in *Selected Readings on Palm Oil and Its Uses*. Kuala Lumpur, Malaysia: Palm Oil Research Institute of Malaysia. p. 11.

Rajanaidu, N., B.S. Jalani, A. Kushairi & V. Rao. 1999. "Oil palm genetic resources-collection, evaluation, utilization and conservation" in *Proceeding of the symposium on the science of oil palm breeding*. N. Rajanaidu & B.S. Jalani (ed.). Bangi, Malaysia: PORIM.

Ravishanker, S. D.K. Kumar, Baranwal, A. Chatterjee & S.S. Solankey. 2013. "Genetic Diversity Based on Clusten and Principal Component Analysis for Yield and Quality Attributes in Ginger (*Zingiber officinale* Roscoe)". *International Journal of Plant breeding and Genetics*. Vol. 7. (3): p. 159-168.

Raychaudhuri, S., J.M. Stuart & R.B. Altman. 2000. "Principal components analysis to summarize microarray experiments: application to sporulation time series". (Paper). *Pacific Symposium on Biocomputing*.

Rokach, L. & O. Maimon. 2010. "Clustering Methods" in *Data Mining and Knowledge Discovery Hand Book*. London: Springer. p. 321-357.

Sammour R., Mustafa, S. Badr, & W. Tahr. 2007. "Genetic Variability of Some Quality Traits in Lathyrus Spp. Germplasm". *Acta Agriculture Slovencia, Genetika*. Vol 43 (1), p.129-140.

Sanni, K.A., I. Fawole, A. Ogunbayo, D. Tia, E.A. Somado, K. Futakuchi, M. Sie, F.E. Nwilene & R.G. Guei. 2010. "Multivariate Analysis of Diversity of Landrace Rice Germplasm". (Paper). *Second African Rice Congress*. Innovation and Partnerships to Realize Africa's Rice Potential. 22-26 March.

Shivani, D., & Ch. Srelakshmi. 2014. "Assessment of Genetic Diversity in Indigenous Germplasm Lines Safflower (*Carthamus tinctorius* L.)". *Canadian Journal of Plant Breeding*. Vol. 2. (1): p. 1-4.

Smiullah, F.A., A. Khan, Afzal, Abdullah, U. Ijaz & R. Iftikher. 2013. "Genetic diversity assessment in sugarcane using principal component analysis (PCA)". *International Journal of Modern Agriculture*. Vol. 2. (1): p. 34-38.

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

Spooner, D., R. van Treuren & M.C. de Vincente. 2005. *Molecular Markers for Genebank Mangament. IPGRI Technical Bulletin No. 10*. Rome, Italy: International Plant Genetic Resources Institute.

Squire, G.R. 1986. "A Physiological Analysis of Oil Palm Trials". *PORIM Buletin*. Vol. 12. p. 12-31.

StatSoft. 2013. Electronic Statistics Textbook. USA. http://www.statsoft.com/Textbook/Cluster-Analysis#general. (Assessed on 3rd January 2014).

Tan, P.N., M. Steinbach & V. Kumar. 2006. *Introduction to Data Mining*. USA: Pearson Addison-Wesley Publisher.

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

Timms, R.E. 1978. "Artefact Peaks in The Preparation And Gas Chromatographic Determination of Methyl Esters". *Australian Journal of Dairy Technology*. March. p. 4-5.

Uhl, N.W. & J. Dransfield. 1987. *Genera Palmarum. A Classification of Palms Based on the Work of Harold E. Moore Jr.* Kansas: The L.H. Bailey Hortorium and the International Palm Society. Allen Press, Lawrence. p.514-516.

Ward, J.H. 1963. "Hierarchical Grouping to Optimize an Objective Function". *Journal of American Statistical Association*. Vol. 58. p. 236-244.

WeldeMicheal G., S. Alamerew, T. Kufa & T. Benti. 2013. "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 (4): p. 47-54.

Wessels Boer, J. G. (1965). The Indigenous Palms of Suriname. palm. Vol. 18, No, 2. P 230-233.

Whitemore, T.C. 1973. *The Palms of Malaya*. Malaysia: Longmans.

Willy, V (2010). "Growth and Production of Oil Palm" in *Soils, Plant Growth and Crop Production VolII*. Encyclopedia of Life Support System.

Wold, S. & M. Sjostrom, 1998. "Chemometrics, Present and Future Success". *Journal of Chemometrics and Intelligent Laboratory Systems*. Vol. 44. p. 3-14.

UNIVERSITI SAINS ISLAM MALAYSIA
ISLAMIC SCIENCE UNIVERSITY OF MALAYSIA

APPENDIX 1

Proximity Matrix of Squared Euclidean Distance

| | Squared Euclidean Distance | | |