

**PRINCIPAL COMPONENT ANALYSIS AND CLUSTERING OF  
EX SITU OIL PALM (*Elaeis guineensis* Jacq.) GERMPLASM**

**Li-Hammed Morufat Abimbola**

(Matric No. 3120239)

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 2014**

**BIODATA OF AUTHOR**

Li-Hammed Morufat Abimbola with passport number A0407566, was born on the 22<sup>nd</sup> August 1988 in the city of Lagos, Nigeria. She is currently residing at PT 5514, Jalan BBN 9/2E, Desa Anggerik, Nilai, Malaysia. She was previously a student of Lagos State University, Ojo (LASU) Lagos, Nigeria (2006-2011) where she obtained Bachelor of Science with Honours in Botany. After her degree, she proceeded after her degree to serve her nation for a year where she tutored as a biology and agricultural science teacher for senior and junior secondary school students for a year. She is at present a Master of Science student of Universiti Sains Islam Malaysia (USIM) (Matric no. 3120239) majoring in Food Biotechnology (plant breeding/genetics options). In her leisure time, she takes to sewing. She can be contacted at [murffylee@gmail.com](mailto:murffylee@gmail.com).



## ACKNOWLEDGEMENTS

My first and foremost appreciation goes to Allah (SWT) who in His infinite mercies has guided me aright, preserved my life to this moment and made this work a reality. I would also like to use this medium to express my heartfelt gratitude to several individuals who were instrumental to my success of completing this research.

To my amiable supervisor, Emeritus Prof. Dr. Jalani Sukaimi for his tutelage and guidance during the period of my research. I say a big thank you to you sir, for your words of encouragement and your confidence in my research.

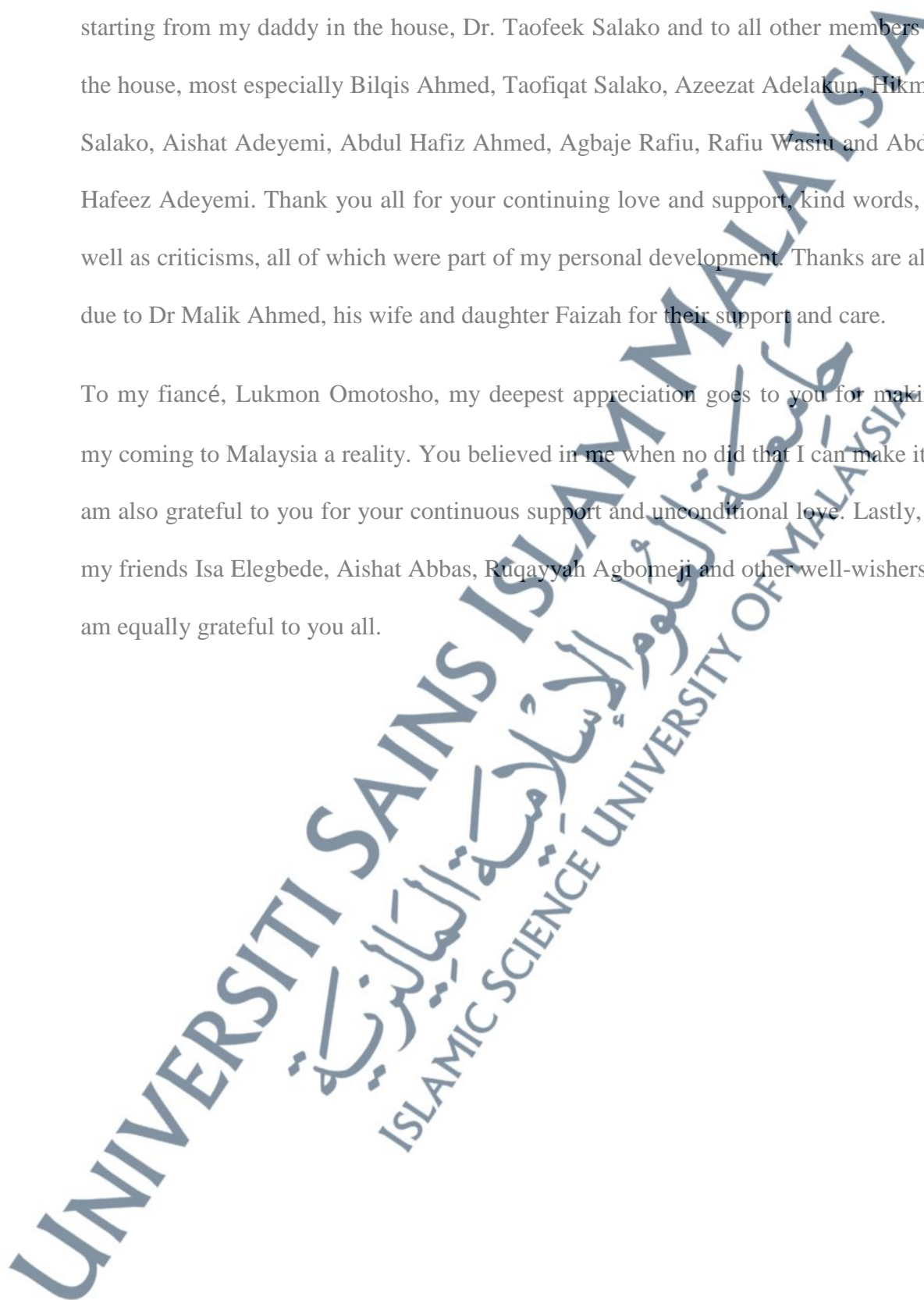
My appreciation is due to Dr. Mohd Sukri. Thank you for being my guide in chemometrics and many invaluable hints that were beneficial to me during the project. Also to the Dean of FST and Head of Biotechnology Program as well as the staff of Food Biotechnology department of USIM, I say a big thank you.

My thanks are due to my other two co-supervisors, Dr. N. Rajanaidu and Dr. Ahmad Kushairi Din at the Malaysian Palm Oil Board (MPOB) for providing the data for this research. I would also like to thank other staff members at MPOB, especially Mr Noh Ahmad for making my research training at Khuang worthwhile.

My thanks would not be adequate without appreciating my mother; Alhaja Sikirat Li-Hammed who has always been my pillar of support. To my late father; Alhaji Mikail Li-Hammed for bringing me to this world and seeing me through my university education and to my sister Fasilat Li-Hammed, for making my sojourn in Malaysia a reality. I pray Allah grant you both Al-Janat Firdaus. Also to all members of my family; especially Aminat Li-Hammed, Ridwan Li-Hammed, Tunde Li-Hammed and Azeez Li-Hammed, I am most grateful.

My thanks would not be complete without acknowledging my Desa Anggerik family, starting from my daddy in the house, Dr. Taofeek Salako and to all other members of the house, most especially Bilqis Ahmed, Taofiqat Salako, Azeezat Adalakun, Hikmot Salako, Aishat Adeyemi, Abdul Hafiz Ahmed, Agbaje Rafiu, Rafiu Wasiu and Abdul Hafeez Adeyemi. Thank you all for your continuing love and support, kind words, as well as criticisms, all of which were part of my personal development. Thanks are also due to Dr Malik Ahmed, his wife and daughter Faizah for their support and care.

To my fiancé, Lukmon Omotosho, my deepest appreciation goes to you for making my coming to Malaysia a reality. You believed in me when no did that I can make it. I am also grateful to you for your continuous support and unconditional love. Lastly, to my friends Isa Elegbede, Aishat Abbas, Ruqayyah Agbomeji and other well-wishers, I am equally grateful to you all.



## ABSTRAK

Pengetahuan kepelbagaian genetik dan hubungan antara bahan germplasma adalah penting untuk pemilihan bahan pembiakbakaan dan juga untuk memudahkan penggunaannya oleh penyelidik biabaka. Oleh itu, penyelidikan ini dijalankan untuk meneroka sejauh mana kepelbagaian dan pertalian ciri-ciri dalam germplasma sawit MPOB-Nigeria menggunakan kaedah statistik yang mudah. Corak kebolehubahan untuk mengenal pasti ciri-ciri yang mengisihkan germplasma telah dikaji dengan menggunakan kaedah PCA dan analisis kelompok. Analisis pekali variasi menunjukkan bahawa kepelbagaian luas wujud di antara BTS dan komponennya. Kajian korelasi pula menunjukkan hubungan yang tinggi dan signifikan di antara BTS dan majoriti daripada ciri komponen kecuali ciri asid lemak. Analisis komponen utama berdasarkan semua ciri yang dikaji menunjukkan bahawa lapan PC pertama mempunyai nilai-eigen lebih besar daripada satu dan menyumbang 90.53% daripada jumlah variasi untuk BTS, ABW, MNW, ME, KF, SF, ODM, OB, KB, OY, TEP, PCS, RL, LL, LW, LN, LA, LAI, BDM, VDM, TDM, e dan f menjadi peranan paling penting pada PC1 yang menyumbang kepada variasi keseluruhan membawa. Analisis kelompok menggunakan kaedah Ward dan SLCA dikumpulkan semua bahan germplasma berdasarkan semua ciri ke dalam lapan kumpulan. Sedikit perbezaan didapati dalam kedua-dua kaedah. Walau bagaimanapun, dari kaedah Ward, Kluster 3 dan 4 mengadungi aksesori dengan ciri hasil yang tinggi, manakala Kluster 2 mengandungi aksesori dengan ciri baik bagi kualiti minyak. Jarak genetik adalah pada aras maksimum dalam Kluster 1 dan Kluster 2 serta juga Kluster 1 dan Kluster 8. Oleh itu, pemilihan ciri bervariasi lebar atau luas boleh digunakan untuk penambahbaikan sawit, manakala penghibridan antara aksesori kelompok yang berbeza dengan kelompok min tinggi bagi ciri terpilih dan jarak maksimum boleh dicapai.

Kata Kunci: Kelapa sawit, germplasma, variasi, korelasi, analisis komponen utama, analisis kelompok.



## ABSTRACT

Knowledge of genetic variability and relations among germplasm materials is crucial for selection of promising breeding materials and also to ease their utilization by plant breeders. Hence, this research was undertaken to explore the extent of variability and associations of traits in the MPOB-Nigerian oil palm germplasm using simple statistical tools. Pattern of variability in order to identify the characters which delineate the germplasm materials was as well studied using principal component analysis and cluster analysis. The results of coefficient of variation indicated that wide variability existed for fresh fruit bunch and its components. Estimates from correlation studies showed high and significant relations between FFB and majority of the traits except the fatty acid traits. PCA based on all the traits studied showed that first eight PCs having eigenvalues greater than one accounted for 90.53% of the total variation with FFB, ABW, MNW, MF, KF, SF, ODM, OB, KB, OY, TEP, PCS, RL, LL, LW, LN, LA, LAI, BDM, VDM, TDM, *e* and *f* being the most important characters on PC1 contributing to the overall variation. Cluster analysis using Ward's method and single linkage method grouped all germplasm materials based on all traits into eight groups. Slight variation was found in both methods. However, from Ward's method Cluster 3 and Cluster 4 contained accessions with high yield traits while Cluster 2 contains accessions with good oil quality traits. Genetic distance was maximum in Cluster 1 and Cluster 2 as well as Cluster 1 and Cluster 8. Thus, selections for characters with high wide variations could be used for oil palm improvement while hybridization between accessions of different clusters with high cluster means for desired traits and maximum distance could be achieved.

Keywords: Oil palm, germplasm, variation, correlation, principal component analysis, cluster analysis.

## CONTENT PAGE

Contents	Page
AUTHOR DECLARATION	i
BIODATA OF AUTHOR	ii
ACKNOWLEDGEMENTS	iii
ABSTRACT	v
ABSTRACT	vi
CONTENT PAGE	vii
LIST OF TABLES	xi
LIST OF FIGURES	xiii
LIST OF APPENDICES	xiv
ABBREVIATION	xv
CHAPTER I: INTRODUCTION	1
1.1 Problem Statement	4
1.2 Objectives	4
1.3 Scope of Research	5
1.4 Organization of Thesis	5
CHAPTER II: LITERATURE REVIEW	6
2.1 Origin, Taxonomy, Botany and Development of Oil Palm	6
2.1.1 Origin	7
2.1.2 Taxonomy	7
2.1.3 Botany	7
2.1.4 Cultivars and Classification	8
2.1.4.1 Classification Based on Fruit Colour	9
2.1.4.2 Classification Based on Forms	9
2.1.5 Oil Palm Breeding and Development	10
2.1.5.1 Oil Palm Breeding in Malaysia	11
2.1.5.2 Oil Palm Breeding in Nigeria	11
2.2 GENETIC RESOURCES	12
2.2.1 Germplasm Collection	12
2.2.2 Germplasm Evaluation	14
2.2.3 Germplasm Utilization	16
2.2.4 Germplasm Conservation	17

2.3 GERMPLASM CHARACTERIZATION AND EVALUATION	18
2.3.1 Genetic Diversity Measurement Tools	18
2.3.1 Grouping Techniques in Measuring Genetic Diversity	21
2.4 MULTIVARIATE STATISTICAL ANALYSIS	22
2.4.1 Principal Component Analysis (PCA)	23
2.4.2 Clustering Analysis	27
2.5 REVIEW OF RESEARCH WORKS ON PRINCIPAL COMPONENT ANALYSIS AND CLUSTERING	30
CHAPTER III: VARIATION AND ASSOCIATION OF TRAITS IN THE MPOB-NIGERIAN OIL PALM GERMPLASM	40
3.1 INTRODUCTION	40
3.2 MATERIALS AND METHODS	41
3.2.1 Plant Breeding Material	41
3.2.2 Data Collection	42
3.2.2.1 Yield and Yield Components	42
3.2.2.2 Bunch Analysis	42
3.2.2.3 Vegetative Measurements and Physiological Traits	45
3.2.2.4 Oil Quality Traits	45
3.2.3 Data Analysis	45
3.3 RESULTS	46
3.3.1 Variation in Traits	46
3.3.1.1 Variations in Yield and Yield Components	46
3.3.1.2 Variations in Brunch/Quality Components	47
3.3.1.3 Variations in Vegetative Traits	48
3.3.1.4 Variations in Physiological Traits	49
3.3.2 Correlation of Traits	49
3.4 DISCUSSION	51
3.5 CONCLUSION	55
CHAPTER IV: PRINCIPAL COMPONENT ANALYSIS OF THE MPOB- NIGERIAN OIL PALM GERMPLASM	63
4.1 INTRODUCTION	63
4.2 MATERIALS AND METHOD	64
4.2.1 Plant Breeding Material	64
4.2.2 Data Collection	65



4.2.3 Standardization	65
4.2.4 Principal Component Analysis (PCA)	66
4.2.5 Theory of PCA	66
4.2.6 Prerequisite Notions and Notations	66
4.2.7 Eigenvalues	67
4.2.8 Scores and Loadings	68
4.2.9 Optimum Number of PCs Extracted	68
4.3 RESULTS	69
4.3.1 PCA of All Variables	69
4.3.2 PCA of Yield and Bunch Quality Components	74
4.3.3 PCA of Vegetative and Physiological Traits	75
4.3.4 PCA of Quality Traits	79
4.4 DISCUSSION	87
4.4.1 Number of Principal Components Extracted	87
4.4.2 Loadings	88
4.4.3 Scores Plot	88
4.4.4 PCA of all Variables	89
4.4.5 PCA of Yield, Yield Components and Bunch Components	92
4.4.6 PCA of Vegetative and Physiological Traits	94
4.4.7 PCA of Oil Quality Traits	96
4.5 CONCLUSION	97
CHAPTER V: CLUSTERING THE MPOB-NIGERIAN OIL PALM	
GERMPLASM	99
5.1 INTRODUCTION	99
5.2 MATERIALS AND METHOD	100
5.2.1 Plants Breeding Material	100
5.2.2 Data Collection	101
5.2.3 Standardization	101
5.2.4 Cluster Analysis	102
5.3 RESULTS	103
5.3.1 Clustering of Oil Palm Populations Based on All Variables	103
5.3.1.1 Single Linkage Clustering (SLCA)	103
5.3.1.2 Ward Hierarchical Clustering (WHC)	104
5.3.2 Clustering Based on Yield and Bunch Components	105
5.3.2.1 Single Linkage Clustering Method	105

5.3.2.2 Ward's Cluster Analysis	106
5.3.3 Clustering Based On Vegetative and Physiological Traits	106
5.3.3.1 Single Linkage Clustering	107
5.3.3.2 Ward's Clustering Method	107
5.3.4 Clustering Based on Fatty Acid Composition	108
5.4 DISCUSSION	109
5.4.1 Clustering Based on All Variables	110
5.4.2 Clustering Based on Yield and Bunch Components	111
5.4.3 Clustering Based on Vegetative and Physiological Traits	112
5.4.4 Clustering Based on Fatty Acid Composition	112
5.4.5 Genetic Distance	113
5.5 CONCLUSION	116
CHAPTER VI: CONCLUSIONS AND RECOMMENDATIONS	127
REFERENCES	129
APPENDICES	142

## LIST OF TABLES

	Page
Table 1: Collection Sites of the MPOB Nigerian Germplasm Materials	44
Table 2: Variation in Yield, Agro-Morphological and Biochemical Traits of the MPOB-Nigerian <i>dura</i> oil palm populations	56
Table 3: MPOB-Nigerian <i>dura</i> Oil Palm Germplasm Means for Yield and Yield Components	57
Table 4: MPOB-Nigerian <i>dura</i> Oil Palm Germplasm Means for Bunch Quality Components	58
Table 5: MPOB-Nigerian <i>dura</i> Oil Palm Germplasm Means for Vegetative Traits	59
Table 6: MPOB-Nigerian <i>dura</i> Oil Palm Germplasm Means for Bunch Quality Components	60
Table 7: MPOB-Nigerian <i>dura</i> Oil Palm Germplasm Means for Oil Quality Traits	61
Table 8: Correlations among Yield, Bunch Quality, Vegetative, Physiological and Oil Quality Traits of the MPOB-Nigerian Oil Palm Germplasm	62
Table 9: Variables Correlation Loading Matrix and Total Variance Explained For All Traits	71
Table 10: Variables Correlation Loading Matrix and Total Variance Explained For Yield and Yield Components	76
Table 11: Variables Correlation Loading Matrix and Total Variance Explained For Vegetative and Physiological Traits	81
Table 12: Variables Correlation Loading Matrix and Total Variance Explained For Fatty Acid Traits	84
Table 13a: SLCA Cluster Means of all Variables in MPOB-Nigerian Oil Palm Germplasm	119
Table 13b: Ward's Cluster Means of all Variables in MPOB-Nigerian Oil Palm	

Germplasm	120
Table 14a: SLCA Cluster Means of Yield and Yield Components in MPOB-Nigerian Oil Palm Germplasm	122
Table 14b: Ward's Cluster Means of Yield and Yield Components in MPOB-Nigerian Oil Palm Germplasm	122
Table 15a: SLCA Cluster Means of Vegetative and Physiological Traits in MPOB-Nigerian Oil Palm Germplasm	124
Table 15b: Ward's Cluster Means of Vegetative and Physiological Traits in MPOB-Nigerian Oil Palm Germplasm	124
Table 16a: SLCA Cluster Means of Fatty Acid Traits in MPOB-Nigerian Oil Palm Germplasm	126
Table 16b: Ward's Cluster Means of Fatty Acid Traits in MPOB-Nigerian Oil Palm Germplasm	126

## LIST OF FIGURES

	Page
Figure 1: Map of Nigeria Showing Collection Sites and Annual Rainfall	43
Figure 2: Principal Component Analysis Plots for All Variables	72
Figure 3: Principal Component Analysis Plots for yield and Bunch Quality Traits	77
Figure 4: Principal Component Analysis Plots for Vegetative and Physiological Traits	82
Figure 5: Principal Component Analysis Plots of for Fatty Acid traits	85
Figure 6: Dendrograms of 44 MPOB-Nigerian Oil Palm Germplasm Based on All 43 Characters of Bunch	118
Figure 7: Dendrograms of 44 MPOB-Nigerian Oil Palm Germplasm Based on 17 Yield and Bunch Quality Traits	121
Figure 8: Dendrograms of 44 MPOB-Nigerian Oil Palm Germplasm Based on 18 Vegetative and Physiological Traits	123
Figure 9: Dendrograms of 44 MPOB-Nigerian Oil Palm Germplasm Based on Eight Fatty Acid Traits	125



## LIST OF APPENDICES

	Page
Appendix A: Procedure for Yield Recording	142
Appendix B1: Bunch Analysis Procedure	143
Appendix B2: Formulae for the Analyses of Bunch Quality Components	144
Appendix C: Formulae for the Analyses of Morpho-Physiological Traits	145
Appendix D: Procedure for Fatty Acid Traits	146
Appendix E1: Proximity Matrix of Euclidean Distance for MPOB-Nigerian Oil Palm Germplasm Based on All Variables	147
Appendix E2: Proximity Matrix of Euclidean Distance for MPOB-Nigerian Oil Palm Germplasm Based on Yield and Yield Components	148
Appendix E3: Proximity Matrix of Euclidean Distance for MPOB-Nigerian Oil Palm Germplasm Based on Morpho-Physiological Traits	149
Appendix E4: Proximity Matrix of Euclidean Distance for MPOB-Nigerian Oil Palm Germplasm Based Fatty Acid Components	150

## ABBREVIATIONS

ABW	average bunch weight
ANOVA	analysis of variance
AVROS	Algemene Vereniging van Rubberplanters ter Oostkust van samantra
BNO	bunch number
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
DIAM	diameter of trunk
<i>e</i>	radiation conversion efficiency
<i>f</i>	fractional interception of radiation
F/B	fruit/bunch
FFB	fresh fruit bunch
FP	frond production
HT	height
IV	iodine value
K/B	kernel/bunch
K/F	kernel/fruit
KY	kernel yield
LA	leaf area
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
NGA	Nigeria
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
S/F	shell/fruit
SLCA	single linkage cluster analysis
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