Journal of Oil Palm Research Vol. 32 (1) March 2020 p. 24-33
DOI: https://doi.org/10.21894/jopr.2020.0016

PRINCIPAL COMPONENT AND CLUSTER ANALYSES ON TANZANIA OIL PALM *Elaeis guineensis* JACQ. GERMPLASM

SUZANA, M*; ZULKIFLI, Y*; MARHALIL, M*; RAJANAIDU, N* and MEILINA ONG-ABDULLAH*

ABSTRACT

Genetic variability in germplasm collection is explored for its role in improving plant breeding strategies. In this study, we assessed the pattern of variation on 15 yield and bunch components in 2191 Elaeis guineensis dura *germplasm from 13 populations (TZA01-TZA13) originating from Tanzania and planted in Malaysia in 1990. The data retrieved from the Malaysian Palm Oil Board-Breeding Information System (MPOB-BIS) were standardised prior to principal component analyses (PCA) and cluster analyses (CA) using SAS 9.4. First four principal component (PC) (PC1-PC4) having eigenvalue >1 accounted for 92.33% of the total variability with values 41.94%, 25.09%, 16.35% and 8.95%, respectively. PC1 has largest positive association with oil to dry mesocarp, oil to bunch and oil to wet mesocarp whereas largest negative association with kernel to fruit, kernel yield and kernel to bunch. PC2 is positively associated with mean fresh fruit bunch, mean average bunch weight, oil and kernel yield whereas negatively associated with shell to fruit and fruit to bunch. CA revealed high genetic variability lies within Tanzania germplasm collection. The combination of PCA and CA is a useful tool to select specific populations to establish core collection for better efficiency in oil palm improvement.*

Keywords: oil palm germplasm, multivariate analysis, eigenvalue.

Date received: 9 July 2018**; Sent for revision:** 9 July 2018**; Accepted:** 20 January 2019.

INTRODUCTION

Oil palm, *Elaeis guineensis* Jacq., a palm of African origin produces palm oil, which has historically been used as a staple food source dating as far back as 5000 years, accounting for 30.8% of today's global production of oils and fat (Mielke, 2018). Its versatility makes it a sought-after ingredient in food and non-food industry. Vast oil palm plantations have been established in tropical countries such as Malaysia and Indonesia due to favourable climatic conditions. In 2017, these two countries produced close to one-third (75.17 million tonnes) of world's total oils and fat production from a planted area of 19.04 million hectares (Kushairi *et al.*, 2018). Malaysia is the world's second largest producer and exporter of palm oil after Indonesia, with a 29% market share of an annual export market worth USD 9.7 billion in 2017 (Workman, 2018). Agricultural products including palm oil need to increase by 70% and net exports of oilseeds and vegetable oils are predicted to triple by 2050 to fulfil market demand as the population growth touching 9 billion (Kushairi *et al.*, 2017).

The introduction of Deli *dura* seeds from four Bogor palms from Indonesia to Malaya (now known as Malaysia) in 1911 and 1912 (Jagoe, 1952) marked the beginning of the Malaysian oil palm industry. Planting materials of *dura* fruit form was planted then until planting materials of *tenera* fruit form with higher oil content was discovered,

* Malaysian Palm Oil Board,
6 Persiaran Institusi, Bandar Baru Bangi,
43000 Kajang, Selangor, Malaysia.
E-mail: suzana.mustaffa@mpob.gov.my

and the monogenic inheritance that governs the variation in fruit forms was reported by Beirnaert and Vanderweyen (1941). This discovery had led to further growth of the oil palm industry worldwide. In order to realise the potential of this industry, planting materials of *tenera*/*pisifera* fruit forms of AVROS lineage were brought into Malaya in 1957 (Kushairi, 2009). Strategies such as this which are imbibed as part of our breeding improvement programmes have helped and placed Malaysia as the world's leading producer and exporter of palm oil in the 1960s (Ooi *et al.*, 2014). However, the genetic base was still considered narrow at that time, thus limiting potential enhancement of fresh fruit bunch (FFB) yield and its components, *e.g*. bunch number and bunch weight (Okwuagwu, 1995). In production of planting materials of *tenera* fruit form, which commonly involves crossing between *dura* and *pisifera* (DxP), Marhalil (2009) reported that improvement for *dura* mother palms is required to increase oil palm yield and quality of the commercial DxP planting materials, as the recurrent utilisation of Delis in many breeding programmes without other origins, might inhibit genetic improvement of bunch yield. To address this problem, one of the efforts undertaken was to broaden the gene pool by collecting *E. guineensis* germplasm materials from its natural habitat in Africa. In 1973 and 2010, the Malaysian Palm Oil Board (MPOB), previously known as the Palm Oil Research Institute of Malaysia (PORIM), led the bioprospection activities to 11 African countries and today the collection formed the largest *ex situ* oil palm conservation programme in the world (Rajanaidu, 1994; Rajanaidu and Jalani, 1994).

In oil palm, agronomic and morphological aspects can directly or indirectly influence oil production. Thus, understanding the genetic variability found amongst the wild oil palm collections is essential to identify and select superior palm before it is being utilised in crossbreeding programme with existing advanced populations. Evgenidis *et al.* (2011) claimed that the assessment of germplasm is the most crucial and should be prioritised. This knowledge can be established through assessment on pedigree, morphology, biochemistry, agronomic performance and molecular level (Mohammadi and Prasanna, 2003). Genetic diversity based on morphological variables of many crops have been successfully evaluated using both principal component analysis (PCA) and cluster analysis (CA). For example, PCA and CA have revealed diversity pattern in tomato (Evgenidis *et al.*, 2011), wheat (Khodadadi *et al.*, 2011; Beheshtizadeh *et al.*, 2013), rice (Maji and Shaibu, 2012), ginger (Ravishanker *et al.*, 2013), white bean (Koij and Saba, 2015) and sugar beet (Danojević *et al.*, 2016). However, application of PCA and CA based on field evaluation in characterising oil palm is still considered limited (Li-Hammed *et al.*, 2016), with the only known published studies conducted by Ekezie (2013), Camillo *et al.* (2014), Suzana *et al.* (2016) and Li-Hammed *et al.* (2015; 2016).

Tanzania is located on the eastern part of Africa. Due to its favourable tropical climate, oil palm grows naturally and its products are traditionally used for food, as building materials and for medicinal purposes (Carrere, 2013). MPOB, in collaboration with the Ministry of Agriculture in Tanzania, has successfully collected germplasm in 1986. In Tanzania, open pollinated bunches were collected from one to seven palms from each collection site, with a total of 13 sites, located near Kigoma along the Lake Tanganyika (Rajanaidu *et al.*, 1992). A total of 42 *dura* bunches were collected with full record of their characteristics: mean bunch weight was 18.4 kg, mean fruit weight was 16.9 g, mean nut weight was 8.9 g, mesocarp to fruit was 46.7%, fruit diameter was 2.7 cm, nut diameter was 2.0 cm and kernel diameter was 1.3 cm (Rajanaidu *et al.*, 1992; 2017). Further evaluation on this *dura* collection revealed the potentiality of having individual palms with high bunch index (BI: >0.6), high vitamin E (>1300 ppm), high carotene (>2000 ppm), low lipase (free fatty acids: 2%-10%) and high protein kernel (>20%). Materials involved has been released to the industry through several MPOB-PORIM Series as PS7 (Junaidah *et al.*, 2004), PS8 (Kushairi *et al.*, 2004), PS11 (Mohd Din *et al.*, 2006), PS13 (Maizura *et al.*, 2008) and PS14 (Noh *et al.*, 2015). Some *dura* materials also showed compact characteristics with height increment < 0.30 m yr^{-1} and rachis length < 5.0 m (Marhalil, 2009). In order to optimally exploit these interesting traits, understanding the genetic diversity is crucial to enable selection of the best parents for development of new breeding materials.

MATERIALS AND METHODS

A total of 2283 *dura* palms derived from open pollinated bunches, with the range of 15 to 61 palms per open pollinated bunch were planted at Bukit Lawiang in MPOB Research Station Kluang, Johor, Malaysia in 1990. Palms were laid out in four replicates using randomised complete block design (RCBD) with 16 palms $plot^{-1}$ $replicate^{-1}$ in a triangular form, with a distance at 9 m apart, giving 148 palms ha^{-1}. This experiment was labelled as Trial 0.256 and each population was indicated as TZA01-13 according to the original collection sites in Tanzania, with the range of one to six families per population. Yield data collection (YR) and bunch analysis (BA) on individual palms were carried out from 1994 until 2000. Bunch quality characters were evaluated using protocol developed by Blaak *et al.* (1963). All data were systematically uploaded into MPOB-Breeding Information System (MPOB-BIS).

Three components for bunch yield and its components were used: fresh fruit bunch (FFB), bunch number (BNO) and average bunch weight (ABW). Meanwhile, the 12 components for bunch quality were: mean fruit weight (MFW), mean nut weight (MNW), mesocarp to fruit (MTF), kernel to fruit (KTF), shell to fruit (STF), oil to dry mesocarp (OTDM), oil to wet mesocarp (OTWM), fruit to bunch (FTB), oil to bunch (OTB), kernel to bunch (KTB), oil yield (OY) and kernel yield (KY).

Total number of *dura* palms analysed were 2191 after removing abnormal, supply and dead palms. Population mean for each component, trial mean and standard deviation were obtained using SAS software ver 9.4. Data was then analysed for PCA after being standardised (mean = 0, standard deviation = 1). Standardisation is needed prior to actual analysis to ensure data normality, change the weight of different variables and remove the effect of measurement units (Cao *et al.,* 1999). In PCA, the correlation matrix between all components was calculated with the respective eigenvalues and scores were plotted with respect to PC1 and PC2. Ward's Minimum Variance based on Euclidean distance was used to construct a phylogenetic relationship in CA using NCSS Statistical Software ver 11.0.12.

RESULTS AND DISCUSSION

Mean Performance of Yield and Bunch Quality Traits

Table 1 shows the mean performance of 13 populations for Tanzania *dura* in Trial 0.256 for yield and bunch quality traits. The grand mean performance for FFB yield was 121.05 kg palm^{-1} yr^{-1}, with mean BNO of 13.08 bunches palm^{-1} yr^{-1} and ABW of 9.39 kg palm^{-1} yr^{-1}. Population TZA04 produced the highest mean FFB of all populations (139.15 kg palm^{-1} yr^{-1}) which was about 14.95% higher than the grand mean FFB of this trial and also 9.09% higher than DxP standard cross (SC) (127.56 kg palm^{-1} yr^{-1}). The result obtained by TZA04 was attributed to high BNO of 14.41 kg palm^{-1} yr^{-1} and balanced ABW (9.76 kg palm^{-1} yr^{-1}). Tanzania *dura* populations tend to have higher BNO and lower ABW, a character which was similar to MPOB Nigerian material reported by Isa *et al*. (2008). Population TZA10 obtained the highest ABW (10.37 kg palm^{-1} yr^{-1}) but was still low compared to DxP (14.21 kg palm^{-1} yr^{-1}).

Mean value for MFW which comprises *dura* mesocarp, shell and kernel in this trial was 9.64 g, comparable to the DxP SC (9.30 g). The highest MFW (10.40 g), MNW (5.23 g) and MTF (49.69%) were all from TZA13. Mean values of other traits such as KTF (13.37%), STF (38.99%), OTDM (76.54%), FTB (63.39%), OTB (13.92%) and KTB (8.44%) are considered satisfactory for *dura* standard as compared to previous analysis on Tanzania *dura* germplasm based on family (Marhalil, 2009) and Nigerian *dura* x Deli *dura* introgression (Noh *et al.,* 2014). In terms of OY, mean value of Tanzania *dura* materials (16.74 kg palm^{-1} yr^{-1}) was still very low compared to DxP SC (28.25 kg palm^{-1} yr^{-1}), considering its high KY (10.20 kg palm^{-1} yr^{-1}).

Correlation Analysis

Correlation analysis for 15 traits studied is shown in *Table 2*. For bunch yield characters, FFB is positively correlated to BNO (r = 0.622) and ABW (0.442), but BNO is negatively correlated to ABW (r = -0.419). According to previous studies, it is proven that FFB is positively correlated to BNO and ABW. However, BNO and ABW are negatively correlated, thus showing an antagonistic effect of the two yield component traits in the performance of oil palms (Marhalil *et al.,* 2009; Djonko *et al.,* 2011). In terms of bunch quality components, the highest significant and positive correlation was observed between OTDM - OTWM (0.987), OTB - OTWM (0.955) and MNW – MFW (0.946). OY was positively correlated with all the other traits except STF. The highest negative correlation was obtained between STF - MTF (-0.898) and all other negative correlations showed either moderate (<-0.700) or low (<0.300) correlation.

Principal Component Analysis (PCA)

Dimension-reduction method of PCA is beneficial in exploring the traits contributing most to variation, thus directly assist the process of plant breeding. PCA studies showed that the first four principal component (PC) have eigenvalue >1.0 which accounted for 92.33% of the total variability reliable (*Figure 1*). Eigenvalues >1.0 has been suggested to be the rule of thumb of how many useful PC having any practical significance and should be retained (Iezonni and Pritts, 1991). The remaining components therefore could be neglected as being not of much significance. A similar study by Li-Hammed *et al.* (2016) also reported six PC having eigenvalue >1, but contributed to the lower total variation (85.52%) among the MPOB-Nigerian oil palm germplasm based on yield and yield components traits.

In this study, PC1 has the eigenvalue of 6.291, which represents 41.94% of the total variation in the data set whereas PC2 and PC3 have eigenvalues of 3.764 and 2.452, respectively with additional variation at 25.09% and 16.35% (*Table 3*). This suggests that PC1, PC2 and PC3 represent the equivalent of six, four and two individual variables respectively. Thus, the genotypes from this PC1 to PC3 will be significant and should be selected for further evaluation in selective breeding programme.

TABLE 1. PERFORMANCE AND MEAN COMPARISON OF 13 TANZANIA GERMPLASM POPULATIONS FOR 15 YIELD AND BUNCH QUALITY COMPONENTS

POP	N	FFB (kg palm^{-1} yr^{-1})	BNO (bunches palm^{-1} yr^{-1})	ABW (kg palm^{-1} yr^{-1})	MFW (g)	MNW (g)	MTF (%)	KTF (%)	STF (%)	OTDM (%)	OTWM (%)	FTB (%)	OTB (%)	KTB (%)	OY (kg palm^{-1} yr^{-1})	KY (kg palm^{-1} yr^{-1})
TZA01	185	119.85	13.22	9.24	10.13	5.12	49.35	13.22	37.43	78.08	48.65	63.13	15.17	8.32	18.17	9.91
TZA02	250	114.29	12.04	9.69	9.48	4.87	48.72	13.03	38.25	77.00	46.75	62.03	14.18	8.04	15.95	9.17
TZA03	266	109.55	11.52	9.56	10.05	5.27	47.65	13.24	39.11	76.77	46.41	64.63	14.30	8.52	15.71	9.21
TZA04	227	139.15	14.41	9.76	8.53	4.62	45.42	14.72	39.86	74.71	43.34	62.99	12.41	9.23	17.25	12.81
TZA05	128	127.02	13.45	9.57	9.26	4.71	48.82	13.52	37.66	76.66	46.55	61.97	14.07	8.37	17.78	10.65
TZA06	188	115.49	13.54	8.65	9.88	5.18	47.54	13.11	39.35	75.95	45.06	65.07	13.97	8.51	16.11	9.79
TZA07	226	118.87	13.47	8.88	9.84	5.29	46.32	13.44	40.24	76.82	46.26	64.73	13.89	8.67	16.44	10.28
TZA08	29	126.77	12.78	10.01	8.91	4.61	48.77	14.55	36.68	75.55	44.48	60.51	12.90	8.93	16.20	11.68
TZA09	100	104.83	13.03	8.47	8.51	4.51	46.92	12.95	40.12	75.85	45.18	62.22	13.17	8.04	13.76	8.42
TZA10	142	120.90	11.71	10.37	10.29	5.34	47.88	13.34	38.78	76.77	45.99	62.97	13.90	8.37	16.81	9.99
TZA11	210	125.75	13.63	9.43	9.70	5.04	47.99	13.27	38.74	76.33	45.43	63.11	13.75	8.34	17.07	10.61
TZA12	133	130.70	14.59	9.03	9.68	5.21	46.22	12.76	41.02	76.09	45.24	62.86	13.06	8.00	16.92	10.37
TZA13	107	128.84	13.20	9.78	10.40	5.23	49.69	13.00	37.31	77.93	48.71	64.37	15.59	8.34	20.03	10.57
DxP SC	46	127.56	9.02	14.21	9.30	1.79	80.25	7.50	12.25	76.53	45.67	60.56	22.18	4.58	28.25	5.88
Trial mean		121.05	13.08	9.39	9.64	5.04	47.65	13.37	38.99	76.54	46.03	63.39	13.92	8.44	16.74	10.20
Std deviation		44.65	4.45	2.26	2.32	1.28	5.46	2.45	4.92	2.95	5.64	5.34	2.73	1.62	6.85	4.25

TABLE 2. CORRELATION ANALYSIS AMONG 15 YIELD AND BUNCH QUALITY COMPONENTS OF 13 TANZANIA GERMPLASM POPULATIONS

	MFFB	MBNO	MABW	MFW	MNW	MTF	KTF	STF	OTDM	OTWM	FTB	OTB	KTB	OY	KY
MFFB	1.000	-	-	-	-	-	-	-	-	-	-	-	-	-	-
MBNO	0.622	1.000	-	-	-	-	-	-	-	-	-	-	-	-	-
MABW	0.442	-0.419	1.000	-	-	-	-	-	-	-	-	-	-	-	-
MFW	-0.104	-0.321	0.175	1.000	-	-	-	-	-	-	-	-	-	-	-
MNW	-0.089	-0.195	0.038	0.930	1.000	-	-	-	-	-	-	-	-	-	-
MTF	-0.127	-0.444	0.363	0.429	0.076	1.000	-	-	-	-	-	-	-	-	-

TABLE 2. CORRELATION ANALYSIS AMONG 15 YIELD AND BUNCH QUALITY COMPONENTS OF 13 TANZANIA GERMPLASM POPULATIONS (continued)

	MFFB	MBNO	MABW	MFW	MNW	MTF	KTF	STF	OTDM	OTWM	FTB	OTB	KTB	OY	KY
KTF	0.527	0.148	0.449	-0.539	-0.513	-0.204	1.000	-	-	-	-	-	-	-	-
STF	-0.110	0.373	-0.560	-0.181	0.157	-0.898	-0.248	1.000	-	-	-	-	-	-	-
OTDM	-0.253	-0.368	0.090	0.774	0.564	0.699	-0.585	-0.429	1.000	-	-	-	-	-	-
OTWM	-0.222	-0.318	0.069	0.702	0.470	0.724	-0.553	-0.468	0.987	1.000	-	-	-	-	-
FTB	-0.183	0.034	-0.344	0.606	0.730	-0.195	-0.387	0.368	0.298	0.272	1.000	-	-	-	-
OTB	-0.241	-0.351	0.072	0.780	0.547	0.739	-0.534	-0.490	0.946	0.955	0.439	1.000	-	-	-
KTB	0.476	0.168	0.322	-0.304	-0.223	-0.287	0.911	-0.125	-0.498	-0.475	0.022	-0.379	1.000	-	-
OPY	0.665	0.265	0.423	0.510	0.346	0.450	0.040	-0.462	0.505	0.542	0.196	0.562	0.123	1.000	-
KPY	0.911	0.520	0.450	-0.284	-0.246	-0.203	0.814	-0.165	-0.448	-0.415	-0.212	-0.394	0.782	0.473	1.000

TABLE 3. CONTRIBUTION OF EACH VARIABLE TO THE EXTRACTED FOUR PCS WITH RESPECTIVE EIGENVALUES AND PROPORTION OF TOTAL VARIABILITY

Traits	PC1	PC2	PC3	PC4
MFFB	-0.168	0.364	0.285	-0.220
MBNO	-0.186	0.017	0.377	-0.555
MABW	0.002	0.393	-0.148	0.337
MFW	0.335	0.083	0.244	0.221
MNW	0.262	-0.020	0.378	0.316
MTF	0.265	0.255	-0.305	-0.143
KTF	-0.294	0.286	-0.071	0.230
STF	-0.130	-0.380	0.334	0.038
OTDM	0.378	0.096	-0.001	-0.095
OTWM	0.366	0.114	-0.017	-0.166
FTB	0.167	-0.125	0.459	0.276
OTB	0.373	0.123	0.031	-0.052
KTB	-0.241	0.254	0.121	0.376
OY	0.143	0.402	0.273	-0.223
KY	-0.246	0.369	0.205	-0.019
Eigenvalue	6.291	3.764	2.452	1.342
Variance (%)	41.94	25.09	16.35	8.95
Cumulative variance (%)	41.94	67.04	83.38	92.33

PC4 had eigenvalue of 1.342 with corresponding variance percentage of 8.95% which symbolises of only one equivalent variable. Thus, this PC should be considered significant if only it has some biological meaning.

All associations were either low to moderate for both positive and negative contributions with no associations observed above 0.600. PC1 has the highest positive and moderate association with OTDM (0.378), OTB (0.373), OTWM (0.366) and MFW (0.335). Low association on PC1 was observed for MTF (0.265), MNW (0.262), FTB (0.167), OY (0.143) and ABW (0.002). Meanwhile, it was negatively associated with KTF (-0.294), KY (-0.246) and KTB (-0.241). PC2 was positively associated with OY (0.402), KY (0.369), ABW (0.393) and FFB (0.364), whereas negatively associated with STF (-0.380) and FTB (-0.125). PC3 was positively associated with FTB (0.459), MNW (0.378), BNO (0.377) and STF (0.334), whereas negatively associated with MTF (-0.305) and ABW (-0.148).

PC1 to PC3 were plotted to observe the relationship among the 13 populations of Tanzanian germplasm (*Figure* 2). TZA13 and TZA01 were positioned at the top right-hand corner of the score plot and showed the most positive along PC1. These might due to high bunch quality performance obtained by both populations such as highest MTF (49.69% and 49.35%), OTDM (77.93% and 78.08%), OTWM (48.71% and 48.65%), OTB (15.59% and 15.17%) and OY (20.03 kg palm^{-1} yr^{-1} and 18.17 kg palm^{-1} yr^{-1}), respectively. STF ratio for

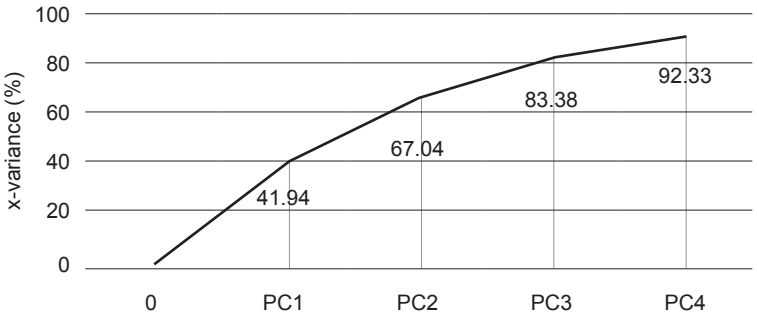

Figure 1. Variance accumulated curve.

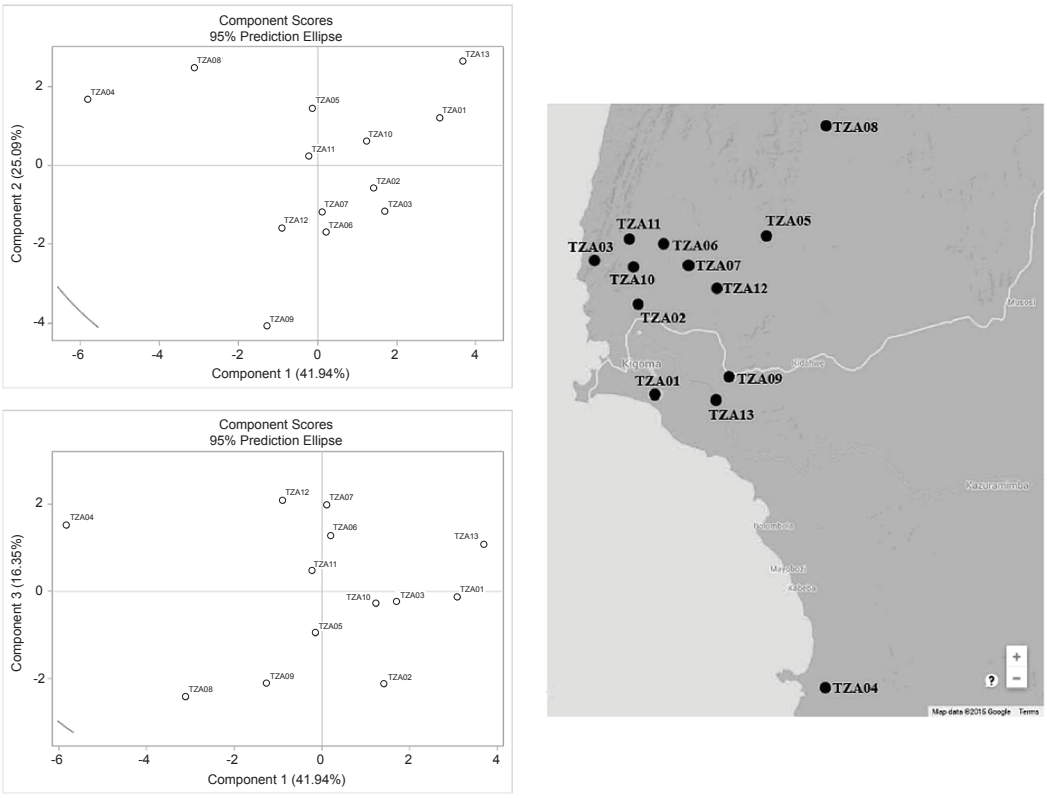

Figure 2. Principal component analyses (PCA) scores extracted from PC1, PC2 and PC3 based on 95% Prediction Ellipse and comparison to the original locations of TZA populations collected in Tanzania.

both populations were also among the lowest (37.43% and 37.31%) and they performed well in yield and its components with their corresponding FFB (128.84 and 119.85 kg palm^{-1} yr^{-1}), BNO (13.20 and 13.22 bunches palm^{-1} yr^{-1}) and ABW (9.78 and 9.24 kg palm^{-1} yr^{-1}) were among the highest compared to the rests. TZA13 was also the most positive in PC2 due to the highest OY (20.03 kg palm^{-1} yr^{-1}) and lowest STF (37.31%). The highest ABW was observed in TZA10 (10.37 kg palm^{-1} yr^{-1}) but it failed to give the highest FFB due to low BNO (11.71 bunches yr^{-1}). High BNO is the easiest way to get high FFB. Nevertheless, the increase in ABW is associated with the reduction in BNO. Thus, when BNO is maximised (*i.e.,* one bunch per leaf), ABW contributes to the difference in the total FFB production. This might due to negative correlation of BNO and ABW which causes great difficulties to maximise both variables in the same palm in order to get higher yield. However, high selection pressure over a long period of time may resolve this obstacle (Kushairi *et al.,* 2003).

TZA04 and TZA08 were positioned at the upper left-hand corner and quite far from the centre (origin) of the plane in the score plot, and these results were in agreement with the original collection sites which were quite distinct from each other. TZA04 and TZA08 were the two most negative along PC1 due to the lowest OTDM (74.71% and 75.55%), OTWM (43.34% and 44.48%), OTB (12.41% and 12.90%), and OTF (304.06% and 321.18%), but highest in KTF (14.72% and 14.55%), KTB (9.23% and 8.93%) and KY (12.81% and 11.68%), respectively. Despite poor performance in bunch quality components, TZA04 performed better in terms of bunch yield with the highest FFB (139.15 kg palm^{-1} yr^{-1}) and among the highest in BNO (14.41 bunches palm^{-1} yr^{-1}) and ABW (9.76 kg palm^{-1} yr^{-1}). This outstanding performance placed TZA04 as the most positive contributors along PC2. Marhalil *et al.* (2009) reported that a family within population TZA04 (TZA04.06) produced the highest FFB yield (148.49 kg palm^{-1} yr^{-1}) and BNO (16.28 bunches palm^{-1} yr^{-1}), compared to DxP standard cross which was about 128.77 kg palm^{-1} yr^{-1} and 9.09 bunches palm^{-1} yr^{-1}, respectively. Thus, it is suspected that this particular family might be the cause of the outstanding FFB of population TZA04 observed in this study.

TZA09 was the most negative along PC2 due to poor performance in some bunch yield and quality components; lowest in FFB (104.83 kg palm^{-1} yr^{-1}), ABW (8.47%), BWT (9.86%), MFW (8.51 g), MNW (4.51 g), OY (13.76 kg palm^{-1} yr^{-1}) and KY (8.42 kg palm^{-1} yr^{-1}). Other populations were positioned quite close to the centre (circled) which indicates that they have average properties of most of the variables. This is also predictable because distance among populations around the centre were not more than 20 km apart and might contribute to the low variability.

Cluster Analysis (CA)

Ward's method is preferred due to its efficiency and widely used for hierarchical clustering analysis (Saracli *et al.,* 2013). In this study, three major clusters were formed in a dendrogram created on the basis of this clustering method based on Euclidean distance (*Figure 3*). The first cluster consisted of TZA01 and TZA13 at the bottom of the dendrogram ($L^2 = 0.371$) and the close proximity in CA reflects similarities in bunch quality traits which were also found to contribute highly to PC1 in PCA. The other branch of this cluster grouped TZA02, TZA03, TZA10, TZA05 and TZA11 together. Only one population from this cluster showed potential characteristic for high ABW (10.37 kg palm^{-1} yr^{-1}), which belonged to TZA10. Second cluster grouped TZA06, TZA07 and TZA12 together in a subcluster branch and TZA09 was positioned as a singleton at another branch, owing to its performance difference for half of the number of components studied. Generally, these two clusters grouped most of the populations having average value for most of the variables possibly due to low relative distance at their geographical locations.

Third cluster containing only TZA04 and TZA08 which were collected from the more southern and northern parts of Tanzania and more isolated from other populations. Even though both populations located approximately 60 km from each other, the same clustering ($L^2 = 1.713$) was influenced by similarities in bunch yield and bunch qualities as discussed in PCA. TZA08 and TZA04 cluster apart from the rest which shows the coherency observed between geographical position and population. It was observed that the geographical distribution patterns of these populations were not random because radial progression of the clusters seems to fit. Similar observations were also reported by Li-Hammed *et al.* (2016) who did the same analysis on MPOB-Nigeria germplasm. Marhalil *et al.* (2009) have suggested a family within TZA04 should be used for further breeding development and also in the formation of core collection due to its outstanding BNO trait.

High level of genetic variability was observed amongst the Tanzanian germplasm materials via the molecular method (Zulkifli *et al.,* 2012) and would be beneficial if it could be exploited to broaden current oil palm genetic base. PCA and CA, combined with more advanced molecular methods, could be used to guide utilisation of MPOB-Tanzania germplasm to its fullest potential for introgression purposes and also regeneration of current germplasm for establishment of a core collection. Core collection was proposed by Frankel (1984) which interpreted as limited set of accessions in an existing collection, with minimum similarities between entries and chosen to represent genetic diversity of a large

collection. Having core collection will reduce the maintenance efforts required against the entire germplasm. Brown (1989) suggested that core collection should be no more than 10% or practically between 5% and 20% of the collections from which they were established. Based on these parameters, the optimum number of Tanzania germplasm that should be preserved is approximately 300 to 600 palms from the whole collection. Both PCA and molecular approach in either way help to determine which populations that should be prioritised for this purpose rather than randomly select the accessions which may be redundant and not bringing any unique value to the next generation. Such systematic approach to germplasm conservation enables us to maximise genetic diversity with minimum germplasm materials for future breeding utilisation.

CONCLUSION

High variability was observed within MPOB-Tanzanian germplasm collection for all 15 traits used in this study. They could be introgressed with advanced breeding materials and may improve selection efficiencies upon some traits that are positively correlated, typically on FFB yield that is of high interest to most breeders. These findings, if coupled with molecular evaluation using reliable markers such as single nucleotide polymorphisms, will provide better insight of the genetic variation of Tanzania oil palm natural population, thus breeding and conservation strategies could be systematically formulated.

ACKNOWLEDGEMENT

The authors wish to express their sincere thanks to the Director-General of MPOB for permission to publish this article.

REFERENCES

Beheshtizadeh, H; Rezaie, A; Rezaie, A and Ghandi, A (2013). Principal component analysis and determination of the selection criteria in bread wheat (*Triticum aestivum* L.) genotypes. *Int. J. Agri. and Crop Sci., 5(18)*: 2024-2027.

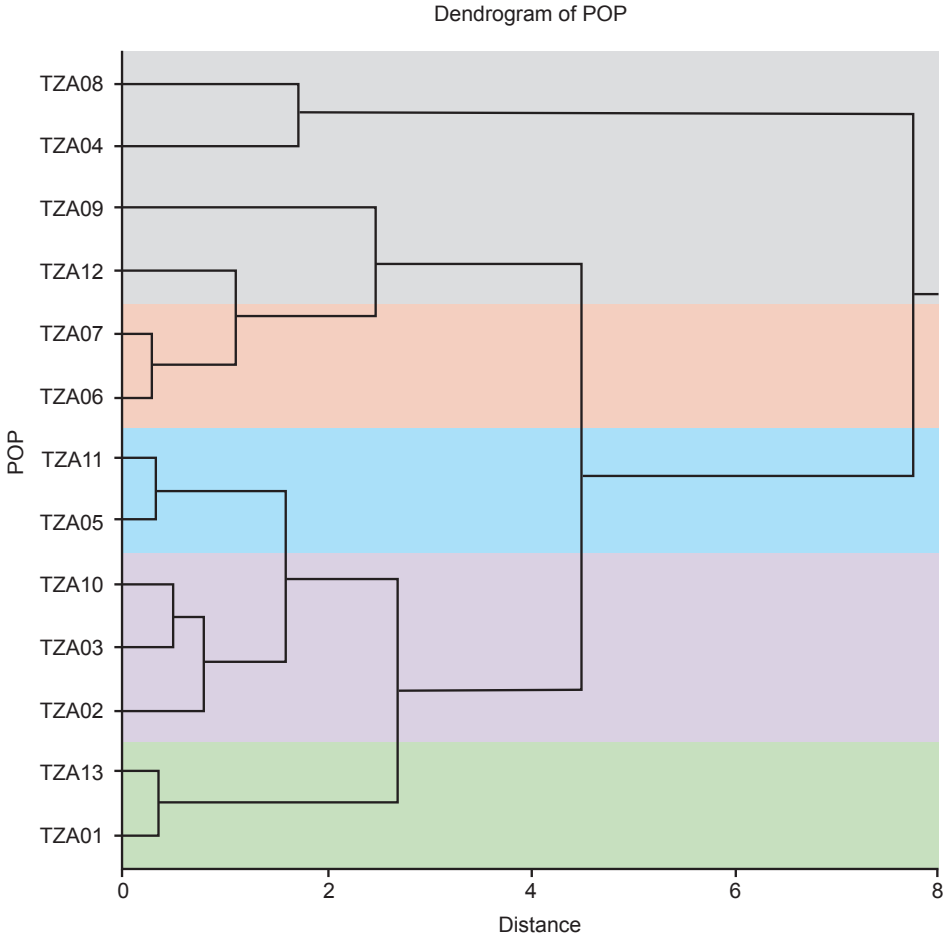

Figure 3. Dendrogram (cluster analysis tree chart) showing the relationship between 13 Tanzanian germplasm populations using 15 diverse characters.

Beirnaert, A and Vanderweyen, R (1941). Contribution à l'étude génétique et biométrique des variétés d'*Elaeis guineensis* Jacquin. *Publ. Inst. Nat. Etude. Agron. Congo Belge. Ser. Sci., 27*: 1-101.

Blaak, G; Sparnaaij, L D and Menendez, T (1963). Breeding and inheritance in the oil palm (*Elaeis guineensis* Jacq.). Part II. Methods of bunch quality analysis. *J. West African Institute Oil Palm Res., 4*: 146-155.

Brown, A H D (1989). The case for core collections. *The Use of Plant Genetic Resources* (Brown, A H D; Frankel, O H; Marshall, D R and Williams, J T eds.). Cambridge University Press, Cambridge, UK. p. 136-156.

Camillo, J; Braga, V; De Abreu Mattos, J K; Lopes, R; Cunha, R N V; Padilha, J and Scherwinski-Pereira, J E (2014). Seed biometric parameters in oil palm accessions from a Brazilian germplasm bank. *Pesq. Agropec. Bras. Brasilia, 49(8)*: 604-612.

Cao, D; William, D D and Williams, N E (1999). Data transformation and standardization in the multivariate analysis of river water quality. *Ecological Appl., 9(2)*: 669-677.

Carrere, R (2013). Oil palm in Africa: Past, present and future scenarios. *WRM Series on Tree Plantation No. 15*. World Rainforest Movement.

Danojević, D; Ćurčić, Ž; Nagl, N; Taški-adjuković, K and Boćanski, J (2016). Evaluation of sugar beet genotypes for root traits by principal component analysis and cluster analysis. *Genetika, 48(1)*: 339-348.

Djonko, H; Kushairi, A; Rajanaidu, N and Jalani, B S (2011). Assessment of bunch yield and its components for Cameroon-based *dura* x *pisifera* oil palm (*Elaeis guineensis* Jacq.) populations. *Proc. of the International Society for Oil Palm Breeders (ISOPB) Seminar*. MPOB, Bangi. p. 167-177.

Ekezie, D D (2013). Principal component analysis, an aid to interpretation data. A case study of oil palm (*Elaeis guineensis* Jacq.). *J. Emerg. Trend Engine. Appl. Sci., 4(2)*: 237-241.

Evgenidis, G; Traka-Mavrona, E and Koutsika-Sotiriou, M (2011). Principal component analysis as a tool in the assessment of tomato hybrids and cultivars. *Intl. J. Agro*. DOI. 10.1155/2011/697879.

Frankel, O H (1984). Genetic perspectives of germplasm conservation. *Genetic Manipulation: Impact on Man and Society* (Arber, W K; Llimensee, K; Peacock, W J and Starlinger, P eds.). Cambridge University Press, Cambridge, UK. p. 161-170.

Iezonni, A P and Pritts, M P (1991). Applications of principal component analysis to horticultural research. *Horticultural Sci., 26*: 334-338.

Isa, Z A; Ong, K P; Norasyikin, I and Suboh, O (2008). Performance of MPOB-Nigerian Population 12 at Kulim - An update. *Proc. of the 3rd Seminar on Performance of PS1 and PS2 Materials and Elite Germplasm*. MPOB, Bangi. 139-149.

Jagoe, R B (1952). Deli oil palms and early introductions of *Elaeis guineensis* to Malaya. *Malay Agri. J., 35*: 4-11.

Junaidah, J; Kushairi, A; ISA, Z A; Mohd Din, A; Noh, A and Rajanaidu, N (2004). PS7: High bunch index breeding population. *MPOB Information Series No. 221*. MPOB, Bangi.

Khodadadi, M; Fotokian, M H and Miransari, M (2011). Genetic diversity of wheat (*Triticum aestivum* L.) genotypes based on cluster and principal component analysis for breeding strategies. *Aust. J. Crop Sci., 5(1)*: 17-24.

Koij, F S and Saba, J (2015). Using cluster analysis and principal component analysis to group lines and determine important traits in white bean. *Procedia Environ. Sci., 29*: 38-40.

Kushairi, A (2009). Role of oil palm breeding in wealth creation and quality of life. *Proc. of the 8th Malaysia Congress on Genetics*. Genting Highlands, Malaysia. p. 14-20.

Kushairi, A; Rajanaidu, N; Mohd Din, A; Isa, Z A; Noh, A and Junaidah, J (2003). PS5: Breeding populations selected for thin shell *tenera*s. *MPOB Information Series No. 183*. MPOB, Bangi.

Kushairi, A; Rajanaidu, N; Sundram, K and Maizura, I (2004). PS8: Vitamin E breeding population. *MPOB Information Series No. 222*. MPOB, Bangi.

Kushairi, A; Singh, R and Ong-Abdullah, M (2017). The oil palm industry in Malaysia: Thriving with transformative technologies *J. Oil Palm Res. Vol. 29(4)*: 431-439.

Kushairi, A; Loh, S K; Azman, I; Elina, H; Meilina, O-A; Zanal Bidin, M N R; Razmah, G; Shamala, S and Ahmad Parveez, G K (2018). Oil palm economic performance in Malaysia and R&D progress in 2017. *J. Oil Palm Res. Vol. 30(2)*: 163-195.

Li-Hammed, M A; Kushairi, A; Rajanaidu, N; Mohd Sukri, H; Che Wan Zanariah, C W N and Jalani, B S (2015). Multivariate analysis of vegetative and physiological traits in oil palm

(*Elaeis guineensis* Jacq.) germplasm. *Expert Opin. Environ. Biol., 4(3)*: 1-5.

Li-Hammed, M A; Kushairi, A; Rajanaidu, N; Mohd Sukri, H; Che Wan Zanariah, C W N and Jalani, B S (2016). Genetic variability for yield, yield components and fatty acid traits in oil palm (*Elaeis guineensis* Jacq.) germplasm using multivariate tools. *Intl. J. Agri. Forest Plant, 2*: 219-226.

Maizura, I; Kushairi, A; Noh, A; Marhalil, M; Wong, Y T and Sambanthamurthi, R (2008). PS13: Breeding populations selected for low lipase. *MPOB Information Series No. 400*. MPOB, Bangi.

Maji, A T and Shaibu, A A (2012). Application of principal component analysis for rice germplasm characterization and evaluation. *J. Plant Breed. Crop Sci., 4*: 87-93.

Marhalil, M (2009). *Performance of MPOB-Tanzania Oil Palm Germplasm Collection*. MPOB 134th VIVA Committee Meeting. MPOB, Bangi.

Mielke, T (2018). *Oil World*. International Statistical Agricultural Information (ISTA) (Falk, S ed.). Mielke GmbH, Hamburg, Germany.

Mohammadi, S A and Prasanna, B M (2003). Analysis of genetic diversity in crop plants-salient statistical tools and considerations. *Crop Sci., 3*: 1235-1248.

Mohd Din, A; Noh, A; Mohd Isa, Z A; Maizura, I; Kushairi, A and Rajanaidu, N (2006). PS11: High carotene *E. guineensis* breeding population. *MPOB Information Series No. 312*. MPOB, Bangi.

Noh, A; Rafii, M Y; Mohd Din, A; Kushairi, A; Norziha, A; Rajanaidu, N; Latif, M A and Malek, M A (2014). Variability and performance evaluation of introgressed Nigerian *dura* x Deli *dura* oil palm progenies. *Genet. Mole. Res., 13(2)*: 2426-2437.

Noh, A; Rajanaidu, N; Kushairi, A; Mohd Din, A and Wan Nooraida, W M (2015). PS14: Oil palm breeding population selected for high kernel. *MPOB Information Series No. 566*. MPOB, Bangi.

Okwuagwu, C O and Tai, G C C (1995). Estimation of variance components and heritability of hunch yield and yield components in the oil palm (*Elaeis guineensis* Jacq.). *Plant Breeding, 114*: 463-465.

Ooi, Z X; Hanafi, I; Azhar, A B and Teoh, Y P (2014). A review on recycling ash derived from *Elaeis guineensis* by-product. *Bioresour., 9*: 7926-7940.

Rajanaidu, N (1994). *PORIM Oil Palm Genebank*. PORIM, Bangi. 19 pp.

Rajanaidu, N and Jalani, B S (1994). Oil palm genetic resources – collection, evaluation, utilisation and conservation. Paper presented at the PORIM Colloquium on Oil Palm Genetic Resources. PORIM, Bangi.

Rajanaidu, N; Jalani, B S; Kushairi, A and Rao, V (1992). Oil palm genetic resources: Collection, evaluation, utilisation and conservation. *Proc. of the Intl. Society of Oil Palm Breeders (ISOPB)*. PORIM, Bangi. p. 219-255.

Rajanaidu, N; Kushairi, A and Mohd Din, A (2017). Evaluation of oil palm genetic resources. *Monograph: Oil Palm Genetic Resources* (Rajanaidu, N ed.). MPOB, Bangi. p. 93-215.

Ravishanker; Kumar, S; Baranwal, D; Chatterjee, A and Solankey, S (2013). Genetic diversity based on cluster and principal component analysis for yield and quality attributes in ginger (*Zingiber officinale* Roscoe). *Intl. J. Plant Breed. Genet., 7(3)*: 159-168.

Saracli, S; Doğan, N and Doğan, I (2013). Comparison of hierarchical cluster analysis methods by cophenetic correlation. *J. Inequalities and Applications, 203*: 1-8.

Suzana, M; Zulkifli, Y; Marhalil, M; Rajanaidu, N and Mohd Din, A (2016). Principal component and cluster analysis as a tool in the assessment of genetic variability of Sierra Leone germplasm populations. *Transactions of Persatuan Genetik Malaysia, 3*: 213-216.

Workman, D (2018). Palm oil exports by country. World's top exports (WTEx). http://www.worldstopexports.com/palm-oil-exports-by-country/

Zulkifli, Y; Maizura, I and Rajinder, S (2012). Genetic diversity study of *Elaeis guineensis* germplasm using EST-SSR. *J. Oil Palm Res. Vol. 24*: 1368-1377.