Determination of Selection Index of Cocoa (*Theobroma cacao* L.) Yield Traits Using Regression Methods

Bayu Setyawan[1*)], Taryono[2)] and Suyadi Mitrowihardjo[2)]

[1)]Indonesian Coffee and Cocoa Research Institute, Jl. PB. Sudirman 90, Jember 68132, Indonesia
[2)]Faculty of Agriculture, Gadjah Mada University, Jl. Socio-Justicia, Bulaksumur, Yogyakarta, Indonesia
[*)]Corresponding author: setyawanbayu45@gmail.com

Abstract

The increasing chocolate consumption has not been followed by growing production of dry cocoa beans. In order to support the increase in cocoa production, planting materials with high yield are needed. The objective of this research was to determine the components of cocoa traits affecting weight of dry cocoa beans, and set a selection index for superior cocoa trees. The experiment material were four cocoa hybrid populations of which their family ancestry were unknown, and were planted on Samigaluh Plantation, Yogyakarta, and Segayung Plantation, Central Java. Observations and data collection were conducted on four plant populations. The observations were undertaken for three years, by observing plant traits, including pod length, pod diameter, husk thickness, cavity diameter, pod fresh weight, cocoa bean/pod fresh weight, husk fresh weight, dry weight of cocoa beans/pod, number of cocoa beans/pod, dry weight per cocoa bean. The collected data were analyzed using path and regression analysis methods. The results showed that pod diameter (X_4), fresh pod weight (X_5), number of cocoa beans/pod (X_8), and dry weight/cocoa bean (X_9) were used to form a selection index resulting the equation $I = 0.0792 X_4 + 0.1330 X_5 + 0.0106 X_8 + 0.1349 X_9$ furthermore will be used in the selection of cocoa trees. Ten cocoa plants from seeds having the highest general selection index were D 034, D 003, D 015, A 054, D 004, D 033, D 041, A 157, D 036, and D 025 will be selected for further evaluation.

Keywords: cocoa, selection index, origin seeds, regression, selection criteria

INTRODUCTION

Chocolate consumption in some countries throughout the world has increased although the cocoa production as a basic ingredient of chocolate remains lower than its consumption (Morris, 2014). The increase in crop production is determined by the use of superior planting materials (Chairuman, 2013). In order to obtain these superior cultivars, plant breeding activities are required. Cocoa breeding activities have produced a number of superior cultivars, which some of those have already been released with quite high yields in the range about 2–3 tons/ha (Puslitkoka, 2015). While the potential of those superior clones is high, Indonesian cocoa production remain low at around 709,331 tons in 2014, with an area approximately of 1,719,087 ha that the productivity of Indonesian cocoa only around 413 kg/ha/year (Ditjenbun, 2014).

Efforts to gain high superior cocoa varieties are constantly done in order to evaluate and get a good method in selecting crossbred plants or genotypes resulted from exploration. Crossing among genotypes is able to improve genetic diversity. Such high diversity becomes a potential to obtain superior plants through selection activities (Van den Hurk, 2009).

Weight of dry cocoa beans is the most crucial component in cocoa cultivation (Zahouli *et al.,* 2010). Cocoa yield is an accumulation of dry weight beans and the number of pod produced by cacao trees. Based on principal component analysis performing that pod index, fresh weight of total cocoa bean, pod length, pod shape, and number of pods are of which cacao traits affecting the consider-able variation of cacao in amount of 33% (Bekele *et al.,* 2008). Selection of several plant traits jointly arranged in a selection index turns out to be 6.68 times more efficient than direct selection to plant yields (Muhe, 2011). According to Sarutayophat (2012), use of path analysis in increasing plant yield, the selection focus on plants traits associated with yields showing more significantly positive results. In index selection, the selection of plant traits as selection criteria is carried out using several methods. Kayan & Adak (2012) used several statistical methods, such as path analysis, regression analysis and multiple regression analysis. Such methods are often employed by considering the economic value possessed by the traits to be chosen, and the economic value depends on multiple traits selection.

Through regression method approach in this research, components of cocoa pod traits affecting the yield weight of dry cocoa beans can be determined. Furthermore, the selection index will be established for plant population of cocoa plants from seeds. In the selection index, potential cacao trees are chosen for the development of superior clones with the highest yield of 1% among the population being tested.

MATERIALS AND METHODS

The research was conducted at peak cacao harvest season, at the second half period, starting from 2011 to 2013 in Segayung Cocoa Plantation of PT. Pagilaran,

Central Java, at an altitude of 50 m asl. (above sea level) with C climate type according to Schmidt-Ferguson. This study was also carried out in plant population at Samigaluh Cocoa Plantation of PT. Pagilaran, Yogyakarta at the altitude of 400 m asl. with climate type B according to Schmidt-Ferguson (Sugiharyanto *et al.*, 2009; Wibisono *et al.*, 2013). Four plant populations of cacao derived from hybrid seeds, planted in 1979 and the precise parents were unknown (free cross-mix seedlings, clones KKM 04, KKM 22, RCC 70, RCC 71, RCC 72 and RCC 73) were planted at a spacing distance of 3 m x 3 m without shade trees. Out of the popula-tion, 300 numbers of cocoa trees were chosen in North Segayung Plantation, Block IIA (code A), 283 numbers of cocoa plants in North Segayung Plantation, Block IIB 2012 (code B) and IIB 2013 (code C) with a land area of 23.25 ha; and 50 numbers in Samigaluh Plantation (code D) with total area of 2.00 ha. For comparison, pods from clones of KKM 04, KKM 22, RCC 70, RCC 71, RCC 72 and RCC 73 were also observed. From each of the observed trees, 3-6 ripe pods were taken to be evaluated of the yield components, including pod length (X_1), pod diameter (X_2), husk thickness (X_3), pod cavity diameter (X_4), pod fresh weight (X_5), cocoa bean fresh weight per pod (X_6), husk fresh weight (X_7), number of cocoa beans per pod (X_8), cocoa beans dry weight per pod (y = representing the cocoa yields) and dry weight per cocoa bean (X_9).

In the selection index, plant traits as selec-tion criteria were done by several methods. Kayan & Adak (2012) used five statistical methods, namely path analysis, regression analysis, multiple regressions analysis, principal components analysis and dendogram analysis for determining the relationship between yield components and plant yields. In the present research, the selection criteria were deter-mined using path analysis and stepwise

regression analysis to determine pod traits which affecting cocoa yields. Path analysis for determining pod traits as selection criteria affecting the yields can be described in equation (Dewey & Lu *cit*. Tadesse *et al.,* 2009):

$$r_{ij} = P_{ij} \sum r_{ik} P_{kj}$$

Where :

r_{ij} = relationship between independent variables and dependent variables measured by the correlation coefficient.

P_{ij} = component of the direct effects of independent variables on dependent variables (path coefficient).

$\sum r_{ik} P_{kj}$ = sum of indirect effect components subjected to independent variables (i) on dependent variables (j) through other independent variables (k).

Meanwhile, additional method for determining pod traits used as selection criteria was forward stepwise regression. The stepwise procedure to achieve the best regression equation was standardized by inserting free variables, one by one, until a satisfactory regression equation was achieved. The insertion sequence was determined using partial correlation coefficient as a measure of the need for independent variables outside the equation to be entered into the equation (Ing & Lai, 2011). Independent variables included in the ideal regression equation were pod traits used as selection criteria.

Selection index to determine the value of cocoa plants used relationship approach between covariance phenotypic and genotypic associated traits. Meanwhile, determination of the selection index used the following equation (Ibrahim *et al.,* 2012):

$$b_1 P_{11} + b_2 P_{12} + \ldots + b_n P_{1n} = g_{1y}$$
$$b_1 P_{12} + b_2 P_{22} + \ldots + b_n P_{2n} = g_{2y}$$
$$\vdots \quad \quad \vdots \quad \quad \vdots \quad \quad \vdots \quad \quad \vdots$$
$$b_1 P_{1n} + b_2 P_{2n} + \ldots + b_n P_{nn} = g_{ny}$$

Where :

P = phenotypic variance/covariance

G = genotypic variance/covariance

P_{11} = phenotypic trait variant predictive value x_1

P_{12} = phenotypic covariance predictive value between x_1 and x_2

g_{1y} = genotypic covariance predictive value between traits x_1 and x_y (y = yields)

The data employed in previous analysis were firstly standardized as the data derived from different populations. The selection index model previously determined was used to determine superior trees from population of cocoa plants from seeds. Following determination of selection index equation, the equation was utilized to select 5% of plants with the highest index and they were grouped based on similarity with related superior clones using principal component analysis. Analysis in this research used Microsoft Office Excel 2007 and SAS 9.1 software portable for Windows®.

RESULTS AND DISCUSSION

Quantitative traits of living organism including plants are those having normal distribution. The expression of these quantitative traits is affected by more than two genes (multiple genes). The agronomic traits, such as yield, are quantitative trait of which

inheritance is affected by multiple genes. In order to improve the effectiveness on the selection of plant traits, a selection should be performed to the several traits associated with the characteristics of yield (Rahman *et al.*, 2007). In multiple traits selection, there are several methods which can to be used, one of which being the use of an index to determine the value of plant (selection index) (Ibrahim *et al.*, 2012). Prior establishing the selection index for cocoa, yield component traits associated with the cocoa yield should be known.

Path analysis showed that the traits can be used as selection criteria of plants traits to increase plant yield, and the criteria selection may either affect directly or indirectly (Senapati *et al.*, 2009). The cocoa yields in this research used dry weight of cocoa beans per pod data. There were two determinations of selection criteria, the first by using path analysis. The path analysis result is shown in Table 1. The results indicate that positive and significant path correlation (r = 0.8069) was found between dry weight per bean and yields and path coefficient value ($PX_9 = 0.7522$) was high. It indicated that the dry weight per cocoa bean had a significant effect in improving yields. When selecting plants having high dry weight per

cocoa bean, the plants with high yields were obtained. The research by Anita-Sari & Susilo (2013) showed that dry weight per cocoa bean had high direct value on dry weight of cocoa beans per pod produced. Similarly, during selection of plants having a large number of cocoa beans per pod, plants with high yield were obtained (r = 0.5908 and $PX_8 = 0.5228$).

Cavity diameter had positive and significant correlation (r = 0.4294) on cocoa yields. The trait of cavity diameter and dry weight per cocoa bean had high indirect effect (0.2583). Traits of fresh weight of pod and dry weight per cocoa bean had indirect and positive effect on increasing yields for having significant and positive correlation (r = 0.5882), and indirect effect (0.3612). The four pod traits above affected the formation of dry cocoa beans. Of the three traits, wet weight per pod, dry weight per bean, and number of beans per pod are associated with cocoa beans. Pod diameter trait affected cocoa beans of pod, as the larger the cacao cavity diameter, the more optimal the development of the pod will be. The research by Chukwudi & Agbo (2014) mentioned that the larger pod cavity diameter, the bigger cocoa bean development it will affect.

Table 1. Direct and indirect effects of yield components on cacao yields

Yield components	Direct and indirect effects									r_{xy}
	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9	
X_1	**0.0040**	-0.0260	0.0108	0.0108	0.0035	0.0170	0.0054	0.1396	0.2265	0.3916
X_2	0.0020	**-0.0514**	0.0142	0.0364	0.0040	0.0183	0.0072	0.1363	0.3511	0.5182
X_3	0.0019	-0.0322	**0.0227**	-0.0067	0.0031	0.0181	-0.0002	0.0358	0.2051	0.2477
X_4	0.0008	-0.0349	-0.0028	**0.0535**	0.0022	0.0065	0.0093	0.1365	0.2583	0.4294
X_5	0.0027	-0.0395	0.0135	0.0224	**0.0052**	0.0257	0.0084	0.1886	0.3612	0.5882
X_6	0.0020	-0.0279	0.0122	0.0103	0.0040	**0.0338**	-0.0088	0.1239	0.2743	0.4237
X_7	0.0008	-0.0130	-0.0001	0.0174	0.0015	-0.0104	**0.0285**	0.1379	0.1374	0.2999
X_8	0.0011	-0.0134	0.0016	0.0140	0.0019	0.0080	0.0075	**0.5228**	0.0474	0.5908
X_9	0.0012	-0.0240	0.0062	0.0184	0.0025	0.0123	0.0052	0.0329	**0.7522**	0.8069

Notes: Figures on yield component section in bold are direct effects, whilst those not in bold are indirect effects: pod length (X_1), pod diameter (X_2), husk thickness (X_3), pod cavity diameter (X_4), pod fresh weight (X_5), cocoa bean fresh weight per pod (X_6), husk fresh weight (X_7), number of cocoa beans per pod (X_8), dry weight per cocoa bean (X_9); Y: dry weight of cocoa beans per pod.

Table 2. Selected traits based on stepwise forward regression with standardized data using combined populations

Stages	Variables included	Ranks of included variables	Partial R^2	R^2 Partial regression equation	C (p)	F stat	Pr > F
1	X$_9$	1	0.6515	0.6515	21,244.3000	7,283.4000	<.0001
2	X$_8$	2	0.2928	0.9443	122.6595	20,493.8000	<.0001
3	X$_5$	3	0.0007	0.9450	74.3930	49.4100	<.0001
4	X$_4$	4	0.0004	0.9454	50.7255	25.3700	<.0001
5	X$_2$	5	0.0003	0.9457	32.1006	20.4900	<.0001
6	X$_3$	6	0.0002	0.9459	16.4452	17.6100	<.0001
7	X$_7$	7	0.0001	0.9460	14.0175	4.4200	0.0356
8	X$_6$	8	0.0001	0.9461	8.5738	7.4400	0.0064

Notes: pod diameter (X$_2$), husk thickness (X$_3$), pod cavity diameter (X$_4$), pod fresh weight (X$_5$), cocoa bean fresh weight per pod (X$_6$), husk fresh weight (X$_7$), number of cocoa beans per pod (X$_8$), dry weight per cocoa bean (X$_9$).

Besides using path analysis, determination of selection criteria can also use stepwise regression analysis to determine the most influential traits on changes in cocoa yields. The stepwise forward regression analysis result showed that all components, except pod length, significantly affected cocoa yields. Ranking or priority level of pod traits affecting yields is determined from the calculated F values of these variables. The greater the value of calculated F, the more significant the effect, will be. In this case dry weight per cocoa bean, provided the most effecting trait among the other pod traits and this variable had the highest coefficient of determination (R^2) of 0.6515 and decreased by decreasing rating of variables in the stepwise regression analysis. The significance of this variable decreased alongside the decreasing rank in affecting plant yields. According to Hezal *cit.* Singh *et al.* (2013), selection index is consisted of several combinations of traits jointly appeared on each method used in determining the selection criteria. In this study, the appearing traits in the selection criteria were pod cavity diameter, fresh cocoa bean weight, number of cocoa beans/pod, and dry weight/ cocoa beans. Therefore, the pod traits were used in the formation of selection index of cocoa yields. By using the corrected selection index, method, from the four populations,

equation of general selection index was obtained (I) as below:

$$I = 0.0792 \, X_4 + 0.1330 \, X_5 + 0.0106 \, X_8 + 0.1349 \, X_9$$

Based on that general index, each index of cocoa plants in the population can be determined. From a population of 913 cocoa trees used, 1% was selected as best cocoa trees based on high index value (Table 3). The selected trees were later classified using principal component analysis to compare them with superior clones in Cocoa Plantation of PT. Pagilaran in order to observe the superiority of selected plant traits compared to traits of existing superior clones.

The selected plants were later subjected to principal components analysis. In the principal components analysis (Figure 1), the first component had Eigen value of 1.6013, and the second component has an Eigen value of 1.2245. According to Smith (2002), the selection of the components used in the principal analysis is based on the largest Eigen value, and two larger Eigen values normally have a value greater than one. Furthermore, Shlens (2009) described the Eigen value was connected with level of data diversity in factors comprising the principal components. The highest Eigen value of the first principal component had the highest part of the data diversity illustrated in grouping

Table 3. The cocoa trees from seeds selected using general selection index

Tree code	Index value	Tree code	Index value	Tree code	Index value
D034	**120.44**	A109	86.58	A281	77.43
D003	**98.33**	A179	86.31	C130	77.27
D015	**97.28**	D032	85.62	A014	76.98
A054	**97.22**	D039	84.97	C098	76.69
D004	**94.74**	C245	82.38	D016	76.47
D033	**92.21**	C019	81.71	C033	76.31
D041	**90.10**	D002	80.94	A101	75.60
A157	**89.96**	A028	80.35	D013	75.27
D036	**89.88**	A065	79.99	C192	74.03
D025	**88.71**	C017	79.78	C186	73.96
A182	88.29	A219	79.54	A212	73.87
A050	88.21	A280	79.31	A048	73.66
A096	87.64	B098	79.00	B005	73.48
D008	87.43	C193	78.50	A231	73.42
D009	87.29	A180	78.44		
D023	87.23	D012	77.52		

Notes: Trees with bold index is unselected ones based on 1% selection of best plants in the populatio.

chart based on principal component analysis. Figure 1 was compiled by the first component having effects of data diversity at 40.0%, and the second component being 30.6%.

Based on principal component analysis presented in Figure 1, the selected cocoa trees based on selection index apparently can be classified into four groups. The group division was based on the quadrant division (Figure 1). Quadrant division starts from quadrant one (quadrant on upper right position) to four (bottom right) in sequent clockwise. The first principal component is a combination of trait value of pod fresh weight and dry weight per bean and the second principal component is a combination of trait values of pod cavity diameter and number of beans per pod.

The first group (the first quadrant) is a group coded with "Z". The group is one with the superiority of all traits used in the selection index by the value of the first component value and a second positive value, comparator clone of RCC 70 was included in this group. The second group (the second quadrant) under the code "Y" is a group with pod cavity diameter and number of superior seeds, but not for pod fresh weight and dry weight per cocoa bean based on the value of first negative principal components, and the second major component has positive value. The third group (third quadrant) coded with "W" comprised of selected trees having average low value of all pod traits used in the selection index compared to individuals in other groups based on the value of principal components, and both are negative in value. Superior clone of RCC 73 was included in this group. The fourth group (quadrant four) is a group of selected plants having superiority of pod fresh weight and high dry weight per cocoa bean, marked with code "X". In this group, the refference clones KKM 04 and RCC 71 were included based on the value of the first positive principal component, whilst the second principal component was negative in value.
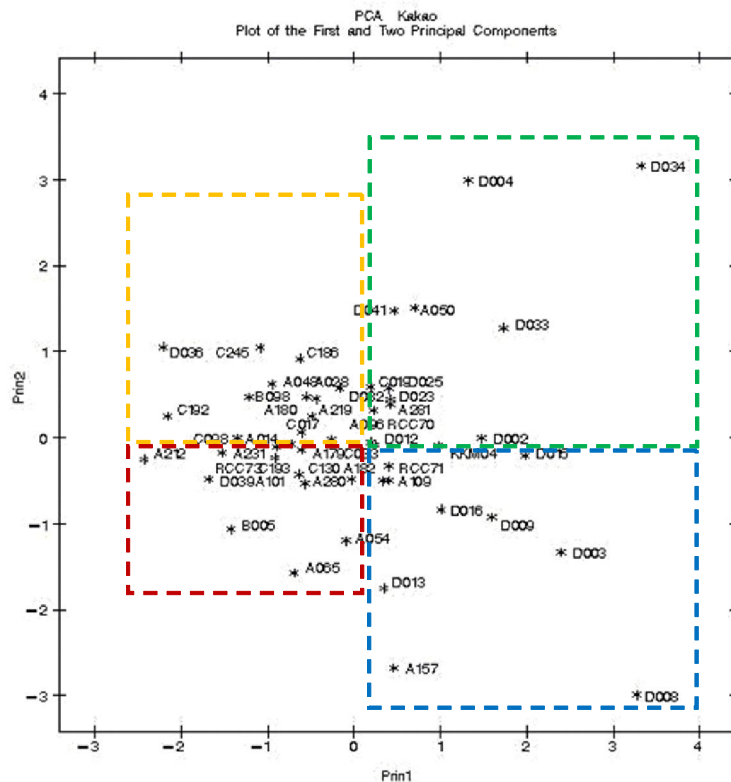
Figure 1. Main component analysis for the 5% of seed derived trees selected with four superior clones (KKM 04, RCC 70, RCC 71 and RCC 73) as control

CONCLUSIONS

Components of pod traits affecting dry weight per pod are pod cavity diameter, pod fresh weight, number of cocoa beans per pod and dry weight per cocoa beans. The equation of general selection index from the population used is $I = 0.0792\ X_4 + 0.1330\ X_5 + 0.0106\ X_8 + 0.1349\ X_9$, with X_4 being the cavity diameter, X_5 being the pod fresh weight, X_8 being the number of cocoa beans per pod, and X_9 is the dry weight per cocoa beans. Ten best cocoa plants derived from hybrid seeds were selected by general those selection index with the codes of D034, D003, D015, A054, D004, D033, D041, A157, D036, and D025.

REFFERENCES

Anita-Sari, I. & A.W. Susilo (2013). Pengembangan kriteria seleksi karakter berat biji pada tanaman kakao (*Theobroma cacao* L.) melalui pendekatan analisis sidik lintas. *Pelita Perkebunan*, 29, 174–181.

Bekele, F.L.; D.R. Butler & G.G. Bidaisee (2008). Upper Amazon Forasteo cacao (*Theobroma cacao* L.) 1 : An assessment of phenotypic relationships in the International Cocoa Genebank, Trinidad. *Tropical Agriculture,* 85, 1–16.

Chairuman, N. (2013). Kajian adaptasi beberapa varietas unggul baru padi sawah berbasis pendekatan pengelolaan tanaman terpadu di dataran tinggi Tapanuli Utara. *Jurnal Pertanian Tropik,* 1, 47–54.

Chukwudi, U.P. & C.U. Agbo (2014). Influence of fruit characteristics on seed and seedling emergence of fluted pumpkin (*Telfairia occidentalis* Hook F.). *The Journal of Animal & Plant Sciences*, 24, 600–605.

Ditjenbun (2014). *Statistik Perkebunan Indonesia : 2013–2015 Kakao*. Direktorat Jenderal Perkebunan, Kementerian Pertanian, Indonesia.

Ibrahim, E.A.; A.W.H. Abdalla; M.E.A. Rahman & A.M. El Naim (2012). Path coefficient and selection indices in sixteen guar (*Cyamopsis tetragonoloba* L.) genotypes under rain-fed. *International Journal of Agriculture and Forestry,* 2, 79–83.

Ing, C.K. & T.L. Lai (2011). A stepwise regression mehod and consistent model selection for high-dimensional sparse linear model. *Statistica Sinica*, 21, 1473–1513.

Kayan, N. & M.S. Adak (2012). Association of some characters with grain yield in chickpea (*Cicerarietinum* L.). *Pakistan Journal of Botany*, 44, 267–272.

Morris, J.A. (2014). *A Taste of the Future : The Trends that Could Transform the Chocolate Industry*. KPMG International, Swiss.

Muhe, K. (2011). Selection index in Durum Wheat (*Triticum turgidum var. Durum*) variety development. *Academic Journal of Plant Sciences*, 4, 77–83.

Puslitkoka (2015). *Pedoman Teknis Budidaya Tanaman : Kakao* (Theobroma cacao *L.*). Pusat Penelitian Kopi dan Kakao Indonesia, Jember.

Rahman, H.; N. Islam; I.H. Khalil; Durrishahwar & A. Rafi (2007). Multiple traits selection in a maize population derived from maize variety Dehqan. *Sarhad Journal Agriculture*, 23, 637–640.

Sarutayophat, T. (2012). Correlation and path coefficient analysis for yield and its components in vegetable soybean. *Songklanakarin Journal of Science and Technology*, 34, 273–277.

Senapati, B.K.; S. Pal; S. Roy; D.K. De & S. Pal (2009). Selection criteria for high yield in early segregating generation of rice (*Oryza sativa* L.) crosses. *Journal of Crop and Weed*, 5, 12–14.

Shlens, J. (2009). *A Tutorial on Principal Component Analysis*. New York University, USA.

Singh, C.M.; G.S. Babu; B. Kumar & S. Mehandi (2013). Analysis of quantitative variation and selection criteria for yield improvement in exotic germplasm of upland rice (*Oryza sativa* L.). *The Bioscan,* 8, 485–492.

Sugiharyanto; M. Nursa'ban & N. Khotimah (2009). *Studi Kerentanan Longsor Lahan di Kecamatan Samigaluh dalam Upaya Mitigasi Bencana Alam*. Laporan Hasil Penelitian Strategis Nasional, Universitas Negeri Yogyakarta, Yogyakarta.

Tadesse, T.; H. Singh & B. Weyessa (2009). Correlation and path coefficient analysis among seed yield and oil content in Ethiopian Linseed germplasm. *International Journal Sustainable Crop Production*, 4, 08–16.

Van den Hurk, A. (2009). The use of genetic resources in plant breeding. *Proceedings of The Second World Seed Conference*, Rome.

Wibisono, I.; Taryono & Nasrullah (2013). Seleksi populasi F2 tanaman kakao (*Theobroma cacao* L.) berdasarkan sifat buah. *Jurnal Cendekia*, 1, 84–99.

Zahouli, G.I.B.; S.T. Guehi; A.M. Fae; L. Ban-Koffi & J.G. Nemlin (2010). Effect of drying on the chemical quality traits of cocoa raw material. *Advance Journal of Food Science and Technology,* 2, 184–190.

0