Original Research Article

Exploring Genetic Variability and Diversity in Exotic Germplasm Collection of Cocoa (*Theobroma cacao* L.)

ABSTRACT

The current research investigates the genetic variation and diversity in exotic accessions of cocoa to identify materials with desired characteristics as well as important traits contributing to yield and variability. The experiment was conducted at Cocoa Research Centre, Kerala Agricultural University, from 2021 to 2023. The study comprises 23 exotic genotypes in which observations were recorded on quantitative traits on flowers, pods, and beans. Statistical analysis was carried out based on the observations recorded. The morphological evaluation revealed ample variability among the genotypes for all the quantitative characters studied. The average dry bean weight recorded was 1.08 grams, which is a desirable trait in selecting individuals with high processing efficiency for the chocolate industry. Characters such as single bean dry weight, pod value, pod index, and efficiency index have recorded high values of GCV, heritability and genetic advance as percent of the mean. Principal component (PC) analysis showed that the first three PCs with more than one Eigen-value contributed to 92.18 percent of total variability among the germplasm. Yield components viz. pod value, total bean weight pod^{-1} and single bean dry weight significantly contribute to maximum variability among the exotic germplasm. Moreover, genotypes were grouped into five clusters based on hierarchical agglomerative clustering technique. Hybridization program involving superior genotypes identified in diverse clusters as parents can produce high-yielding hybrids. The high values recorded for the genetic parameters by yield components suggest the selection of genotypes based on which can result in better-performing progeny.

Keywords: Quantitative traits, Correlation, Principal component analysis, Dendrogram, Pod value

UNDER PEER REVIEW

1. INTRODUCTION

Cocoa (*Theobroma cacao* L.) is a popular beverage crop enjoyed for its bitter beans that serve as the primary raw material for all choco-related confectioneries. It is the most important tree crop in the Malvaceae family and the only source of chocolate. Cocoa originated in the Amazon River basin, which was later referred to as the primary centre of diversity [1]. Over time, the crop spread from Central America to other regions via invaders and traders. Presently, cocoa is grown in Africa, Latin America, south-eastern Asia, and Europe. However, most of the cultivation is confined to the equator due to the specific climatic requirements of the crop. The tropical climate, along with high rainfall and short dry spells in these areas, contribute positively to crop growth. In recent years, the crop has gained the status of a major crop in all growing domains because of its steady demand and reasonable price. In India, cocoa has been mainly cultivated in the southern states since its introduction in late 1960. Currently, the crop covers an area of 1,09,300 hectares, with a production and productivity of 29,800 MT and 0.30 MT/ha, respectively [2]. This perennial crop is having a lifespan of 35 years and has become part of a multi-layered cropping system in major regions, often grown alongside crops like coconut, arecanut, oil palm, and rubber, providing additional income for farmers.

In introduced crops, exotic germplasm serves as the primary source of allelic variation, allowing breeders to access genotypes from different geographical regions. The incorporation of exotic germplasm into breeding programme enhances the genetic base and offers a diverse pool of alleles essential for improving traits such as pest and disease resistance along with yield. The unique agro-climatic conditions in India, necessitate the need of rigorous screening of the available exotic genetic resources to identify candidate genotypes. However, the lack of detailed agronomic information on germplasm diversity hampers their efficient use. Despite progress in molecular techniques, morphological characterization remains essential for exploiting the genetic potential, as it indirectly assesses the agronomic properties of the crop [3]. Genetic variability analysis helps to identify plant material with desired traits and select the best accessions for improvement program. Cacao genetic variability can be understood and utilized through phenotypic characterization. Various quantitative characters including flower diameter, ovary breadth, number of beans pod^{-1}, total bean weight pod^{-1}, single bean dry weight, pod value, pod index and efficiency index were recorded in the study as they reflect the field performance.

Multivariate statistical methods are employed to evaluate genetic diversity and family relationships among different genotypes [4]. Among them, principal component analysis (PCA) helps organize various variables into key unrelated factors [5]. It is also helpful in grouping germplasm and identifying major contributors to genetic diversity [6]. This approach facilitates a better understanding of genetic potential for selection and breeding. In this context, this study aims to evaluate the genetic background of exotic genotypes based on quantitative traits, with the goal of identifying superior genetic stocks for future breeding activities.

2. MATERIALS AND METHODS

The study was conducted at the Cocoa Research Centre, Kerala Agricultural University. Twenty-three exotic genotypes (table 1) collected from the University of Reading, UK were used in this study. The experiment was laid out in completely randomized design with five replications during 2021-23. Pods and fresh flowers were collected from steady bearing cacao trees, and observations were taken. In this study, eight quantitative characters such as flower diameter, ovary breadth, number of beans pod^{-1}, total bean weight pod^{-1}, single bean dry weight, pod value, pod index and efficiency index were taken into account.

UNDER PEER REVIEW

Table 1. List of genotypes

Sl. No.	Genotype
G1	POUND 7/B
G2	CRU 111
G3	ICS 5
G4	GV 5.1
G5	GV 145 UK
G6	ICS 40
G7	TRD 44
G8	SC 20
G9	COCA 3308/A (CHA)
G10	GEBP 914/AF (ADR)
G11	GEBP 617/AF (ADR)
G12	GEBP 180/AM (ADI)
G13	BE 3
G14	CC 11
G15	EQX 69

G16	GU 114/P
G17	PBC 123
G18	MAN 15-2
G19	MAN 15-60
G20	GU 2691/V
G21	PA 56
G22	SC 1
G23	UF 221

Fresh flowers were collected during morning hours and measurements on flower diameter and ovary breadth were made using the stereo microscope model Leica EZ4. For pod characterization, ten ripened pods (two per replication) were collected and their average values were used for statistical analysis. The average value of twenty peeled and dried beans to moisture content below eight per cent was taken as the single bean dry weight. The economic parameters such as pod value, pod index, and efficiency index were calculated as per the standard formula. Pod value is the total weight of dried beans obtained pod^{-1} [7]. It is derived by multiplying single bean dry weight with number of beans pod^{-1}. The pod index estimates the number of pods required to get one kilogram of dry beans [8], whereas the efficiency index indicates the pod weight required to get one gram of dry beans [9]. Ranking of genotypes was carried out by scoring based on eight yield attributing characters. Computation of genetic parameters, correlation analysis and PCA were carried out using the grapesAgri1 package of R software [10]. A dendrogram was constructed using hierarchical agglomerative clustering technique in R software.

UNDER PEER REVIEW

3. RESULTS AND DISCUSSION

Information about variability and genetic relatedness among elite breeding materials is a major component of plant breeding program [11]. In the study, the genetic analysis of the exotic cocoa germplasm based on quantitative characters revealed the presence of ample variability among the genotypes (table 2). The genotypes exhibited the highest significant variation for all the traits assessed in the study (table 3). The distribution of single dry bean weight showed that more than half of the genotypes studied had weight above the minimum standard of 1.00 grams [12] and, with the highest of 1.86g. Number of beans pod^{-1} is an important factor in tracing variability as well as a determinant of crop yield [13]. Among the genotypes, the trait ranged from 26.10 to 48.80, and the lowest and highest values were recorded by genotypes PA 56 and MAN 15-2, respectively. Pod value recorded a mean value of 41.26g. The maximum pod value was recorded for TRD 44 (64.31g), while the minimum was for GU 2691/V (23.42g). Another economic parameter, the pod index, which gives an indirect measurement of yield, ranged between 15.62 and 42.72. Among the 23 genotypes, the efficiency index recorded a mean value 12.25. Ranking of genotypes based on scores obtained for the eight morphological characters, UF 221 and CC 11 ranked first, followed by ICS 40. Meanwhile, PA56 and GU 2691/V scored inferior among the studied germplasm.

Table 2. Mean performance of genotypes for yield contributing characters

Genotype	FD	OB	NB	TBWT	SBDWT	PV	PI	E
G1	11.79	1.19	44.20	130.91	0.86	38.01	26.35	11.36
G2	12.05	0.93	37.80	141.64	1.60	60.68	16.59	7.32

G3	14.90	0.88	40.20	150.82	1.44	57.93	17.31	9.93
G4	12.64	0.98	41.90	100.25	0.89	37.21	26.92	10.83
G5	10.45	1.01	44.80	107.12	0.91	40.91	24.71	9.64
G6	11.89	0.90	40.10	146.90	1.61	64.31	15.62	8.83
G7	9.37	1.03	38.20	110.94	0.80	30.39	33.02	12.26
G8	13.70	0.99	43.10	141.82	1.18	50.78	19.78	10.97
G9	12.14	0.96	29.60	92.34	1.09	32.16	31.19	16.58
G10	12.83	1.27	43.90	124.66	0.88	38.56	26.04	16.57
G11	10.09	0.90	38.30	81.55	0.84	32.32	31.01	9.77
G12	9.82	0.96	47.60	98.97	0.78	37.00	27.11	9.56
G13	10.37	1.13	48.20	141.97	1.12	53.93	18.72	10.75
G14	15.13	0.99	43.70	151.67	1.18	51.09	19.70	9.09
G15	15.57	1.16	36.15	118.11	1.20	43.25	23.20	11.48
G16	10.66	1.00	42.90	95.08	0.71	30.49	33.13	16.73
G17	12.29	1.01	42.10	117.41	1.23	51.77	19.36	9.84
G18	12.70	1.10	49.80	112.92	0.81	40.27	25.23	11.85
G19	12.86	0.94	43.60	119.92	1.04	45.30	22.55	11.83
G20	12.51	0.96	28.40	72.63	0.83	23.42	42.72	15.45
G21	10.95	0.82	27.30	84.37	0.92	25.04	40.08	18.59
G22	11.57	0.90	37.20	119.11	1.28	47.46	21.10	11.84
G23	15.56	1.00	35.90	143.04	1.58	56.64	17.76	8.75

Table 3. ANOVA table for yield contributing characters

Sources of variation	df	FD	OB	NB	TBWT	SBDWT	PV	PI	E
Genotypes	22	16.22	0.06	177.97	2763.37	0.39	659.62	267.74	45.25
Error	92	0.04	0.01	8.64	118.29	0.004	14.30	4.30	3.10
F value		459.61**	5.16**	20.60**	23.36**	100.74**	46.13**	62.12**	14.61**
CV		1.53	10.52	7.31	9.25	5.36	8.80	8.24	15.00
CD (0.01)		0.31	0.18	4.89	18.09	0.10	6.29	3.45	2.93

*** significant at 1% level of significance*

FD – flower diameter, OB – ovary breadth, NB – number of beans pod^{-1}, TBWT – total bean weight pod^{-1}, SBDWT – single bean dry weight, PV – pod value, PI – pod index, E – efficiency index

Validation of genetic parameters for the selected traits will aid in the selection of parents for further breeding activities as it ensures genetic gain. Among the genetic parameters, the values of PCV and GCV were high (>20%) for single bean dry weight, pod value, pod index, and efficiency index. In terms of broad sense heritability, all characters exhibited high values except ovary breadth (table 4). Similar results were reported earlier [14]. High values of GCV and heritability indicates less influence of external environment in their expression. However, the trait ovary breadth showed a moderate heritability. Single bean dry weight, pod value, pod index, and efficiency index showed high PCV, GCV, heritability, and genetic advance in the studied germplasm, indicating additive gene action and hence can be used as a selection criterion for future breeding program.

Table 4. Genetic parameters of yield contributing characters

UNDER PEER REVIEW

Character	Mean	Range	PCV	GCV	H_b^2 *(%)*	GAM
Flower diameter (mm)	12.25	9.37-15.57	14.76	14.68	98.90	30.08
Ovary breadth (mm)	1.00	0.82-1.27	14.22	9.60	45.50	13.34
Number of beans pod^{-1}	40.22	27.30-49.80	18.50	15.44	69.70	26.56
Total bean weight pod^{-1} (g)	117.57	72.63-151.67	21.64	19.56	81.70	36.43
Single bean dry weight (g)	1.08	0.71-1.61	26.32	25.65	95.00	51.51
Pod value (g)	41.26	23.42-64.31	27.97	25.43	82.60	47.61
Pod index	26.30	15.62-42.72	30.62	28.42	86.10	54.34
Efficiency index (%)	12.25	7.32-18.59	29.68	24.62	68.80	42.07

The number of seeds pod^{-1} is primarily influenced by the number of ovules ovary^{-1}, the fertility of the ovules, and the compatible nature of the tree [15]. In the correlation analysis (figure 1), the coefficient values of more than 0.20 are significant at one per cent level of significance. The analysis revealed a positive correlation of ovary breadth (0.28) with the number of beans pod^{-1}. Similarly, the flower diameter exhibited a moderate positive correlation with total bean weight pod^{-1} (0.45) as well as single bean dry weight (0.50) [16]. The positive association, along with high heritability and GAM of these characters, indicates that the selection of genotypes having larger flowers for breeding program brings a high yield. In the correlation analysis, the trait pod value showed a very strong (-0.95) and strong (-0.71) negative correlation with the pod and efficiency indexes, respectively. This association indicates that, similar to the pod index, the low value of the efficiency index is a preferred characteristic for obtaining a better yield. Selection of genotypes were performed earlier [17] based on inter-character associations for good yield.

UNDER PEER REVIEW

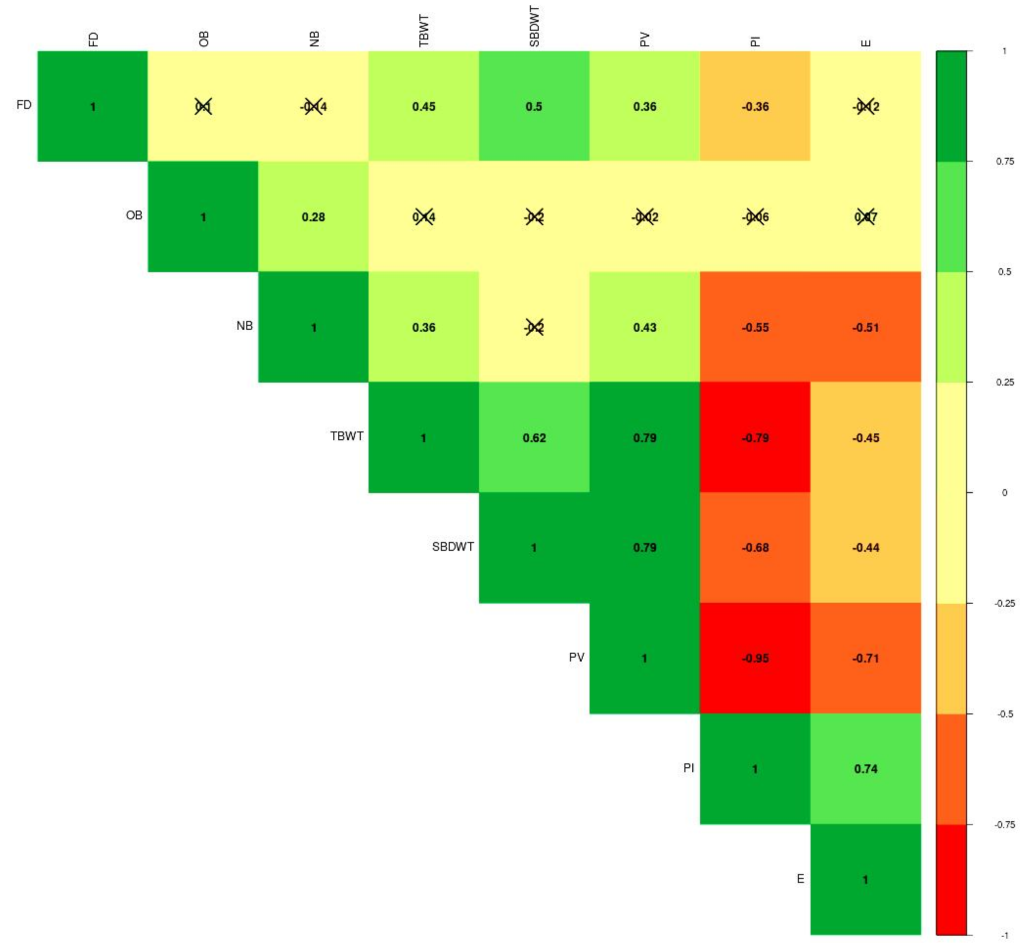

Fig. 1. Correlation of yield attributing components

FD – flower diameter, OB – ovary breadth, NB – number of beans pod^{-1}, TBWT – total bean weight pod^{-1}, SBDWT – single bean dry weight, PV – pod value, PI – pod index, E – efficiency index

Principal component analysis (PCA) was carried out using eight quantitative characters. In PCA, eigenvalues reflect the amount of variation explained by each principal component (PC). The eigenvectors for each trait are presented in table 5. The first three axes explained 90.28 per cent of the total variation across the eight traits used to describe the germplasm, with eigenvalues of 4.48, 1.68, and 1.06, respectively. The first PC alone accounted for 55.99 per cent of the total variation, highlighting its importance in distinguishing between exotic germplasm. Hence, the traits loaded on it play significant roles and are the most important in describing the variation among the cacao clones since PCA measures axes along which variation between genotypes is maximized [18].

In this study, the first principal component was primarily loaded by pod value (0.46), total bean weight per pod (0.44), and single bean dry weight (0.39), while the efficiency index (-0.37) had a negative effect, hence reducing diversity. The second PC was influenced mainly by the number of beans per pod (0.65) and ovary breadth (0.59). Meanwhile, the third PC was mainly loaded by flower diameter (0.68). Key yield related traits, such as pod value, total bean weight pod^{-1}, and single bean dry weight, were contributing to maximum variation among the genotypes, as they were highly loaded in the first principal component. A similar trend in variability by pod value was already reported [19].

The configuration of genotypes and variables along the first two axes, which accounted for 76.99 per cent of the total variation, is plotted in figure 2. The distribution of variables in the biplot indicates the relationships among them. The acute angle between ovary breadth and number of beans pod^{-1}; total bean weight pod^{-1}, pod value, flower diameter and single bean dry weight indicating the positive correlation between the characters. The diagram also depicts the negative correlation of pod index and economic index with the other traits. Moreover, the obtuse angle between ovary breadth and number of beans pod^{-1} with single bean dry weight in biplot demonstrating the chance of obtaining small seeds as the number of beans increases. In this study, total variability of the exotic germplasm is highly influenced by the trait, single bean dry weight. Meanwhile, the configuration of genotypes along the first and second PCs shows the clear separation, allowing grouping of genotypes based on their positions in the PCA biplot [20].

Table 5. Eigen values, loading of variables on principal components and percentage contribution of variance by principal components

Variables	PC1	PC2	PC3	PC 4
FD	0.25	-0.21	0.68	0.62
OB	0.03	0.59	0.54	-0.22
NB	0.19	0.65	-0.23	0.18
TBWT	0.44	0.08	0.14	-0.31
SBDWT	0.39	-0.40	0.04	-0.30
PV	0.46	-0.06	-0.08	-0.21
PI	-0.46	-0.08	0.07	0.09
E	-0.37	-0.07	0.41	-0.55
Eigen value	4.48	1.68	1.06	0.38
Percentage contribution to variance	55.99	20.99	13.29	4.74
Cumulative percentage of variance	55.99	76.99	90.28	95.01

FD – flower diameter, OB – ovary breadth, NB – number of beans pod^{-1}, TBWT – total bean weight pod^{-1}, SBDWT – single bean dry weight, PV – pod value, PI – pod index, E – efficiency index

UNDER PEER REVIEW

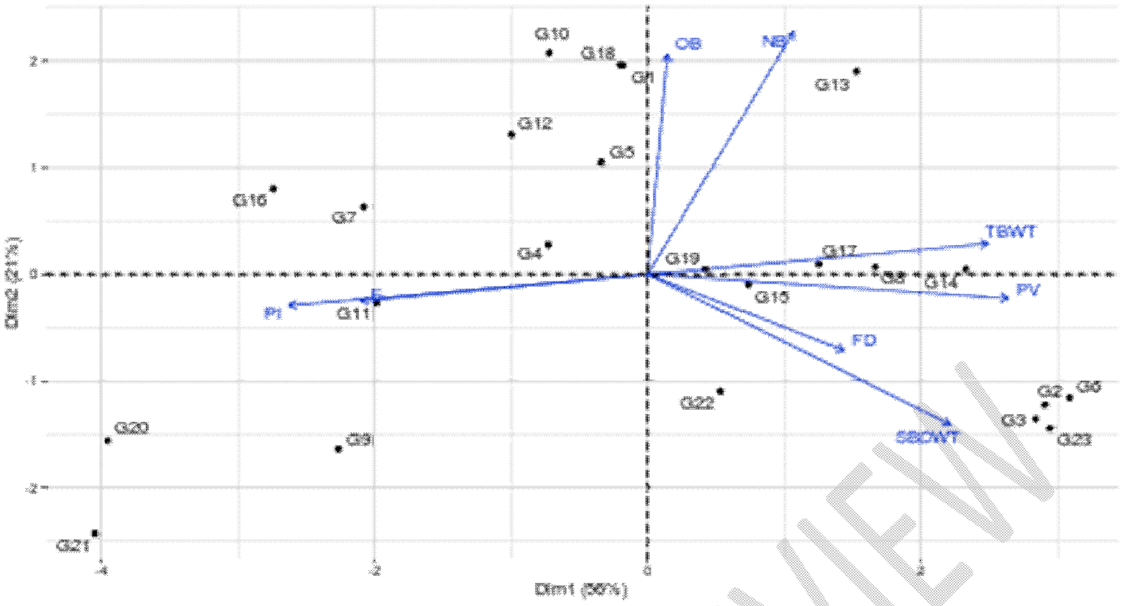

Fig. 2. PCA biplot

PREVIEW

Hierarchical clustering based on the eight variables under study demonstrates a better understanding of the diversity among genotypes. Based on the dendrogram (figure 3), the genotypes form five different clusters. cluster 1 and 4 form the largest clusters including six genotypes each. According to the study, cluster 1 and 2 are more related to one another and are mostly influenced by total bean weight pod^{-1}, pod value, flower diameter and single bean dry weight, while cluster 4 and 5 are related. The three genotypes G9, G20 and G21 are distinct from the rest forming a different cluster. Utilization of the genotypes belongs to most diverse clusters can bring genetic gain by better utilization of heterosis.

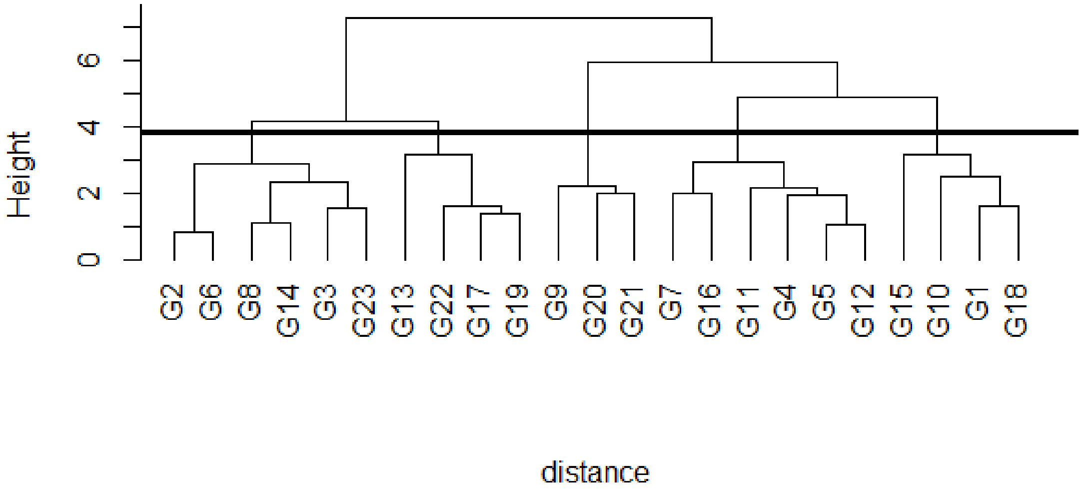

Fig. 3. Cluster analysis of genotypes based on yield contributing characters

4. CONCLUSION

In short, the results revealed significant genetic variation in the germplasm regarding yield-related traits. The presence of distinct variability among germplasm ensures the evolutionary survival of cocoa and provides opportunities for enhancing the crop by focusing on traits exhibiting significant variation. In the present study, among the 23 genotypes evaluated, UF221 and CC11 performed better based on the yield attributes and can be exploited further in crop improvement program. Based on the values of genetic parameters and correlation, characters such as pod value, pod index, efficiency index and single bean dry weight should give emphasis while considering these genotypes for developing genetic stocks. Moreover, among these four, pod value and single bean dry weight have the major contribution to the total variability of this exotic germplasm material, which was evident from the principal component analysis and hence can be fixed as the selection criteria.

Disclaimer (Artificial intelligence)

Author(s) hereby declare that NO generative AI technologies such as Large Language Models (ChatGPT, COPILOT, etc.) and text-to-image generators have been used during the writing or editing of this manuscript.

UNDER PEER REVIEW

REFERENCES

1. Wood GAR, Lass RA. Cocoa. 4th ed. Englond: Longman Group Limited, Longman house; 1985.

2. Indiastat. Area, production and productivity of cocoa in India. Accessed 25 July 2024. Available: https://www.indiastat.com/table/agriculture/area-production-productivity-cocoa-india-1983-1984/371492.

3. Widyasary RA, Susandarini R. Morphological variability and taxonomic affinity of cocoa (*Theobroma cacao* L.) clones from Central Sulawesi, Indonesia. Curr Bot. 2020:11:60-64. https://doi.org/10.25081/cb.2020.v11.5986

4. Santhy V, Rathinavel K, Saravanan M, Meshram M, Priyadharshini C. Genetic diversity assessment of extant cotton varieties based on Principal Component Analysis (PCA) and cluster analysis of enlisted DUS traits. Elec Plant Breed. 2020:11(2):430-438. https://doi.org/10.37992/2020.1102.075

5. Otwe EP, Agyirifo D, Galyuon IKA, Heslop-Harrison P. Molecular diversity in some Ghanaian cowpea [*Vigna unguiculata* L. (Walp)] accessions. Trop Plant Biol. 2017:10(2). https://doi.org/10.1007/s12042-017-9184-9

6. Mohammadi SA. Statistical methods in genetics. Paper presented at the 6th International Statistics Conference, University of Tarbiatmodares, Iran, 26-28 August 2002.

7. Toxopeus H, Jacob VJ. Studies on pod and bean values of *Theobroma cacao* L. in Nigeria. II. number of beans per pod, with special reference to the natural pollination process. Neth J agric Sci. 1970:18:188-94. https://doi.org/10.18174/njas.v18i3.17342

8. Morera J, Mora A, Paredes A. Characterization of a population of ‘Nacional’ cocoa at CATIE, Costa Rica. Turrialba. 1991:41(4):583-88.

9. Jacob VJ, Atanda OA. Pod value studies of Amelonado and Amazon Cacao. Turrialba. 1973:23(3):347-51.

10. Gopinath PP, Parsad R, Joseph B, Adarsh VS. grapesAgri1: collection of shiny apps for data analysis in agriculture. J Open Source Softw. 2021:6(63):3437. https://doi.org/10.21105/joss.03437

11. Mukhtar MS, Rahman M, Zafar Y. Assessment of genetic diversity among wheat (*Triticum aestivum* L.) cultivars from a range of localities across Pakistan using random amplified polymorphic DNA (RAPD) analysis. Euphytica. 2002:128:417-25. https://doi.org/10.1023/A:1021261811454

12. Amma SP, Nair RV, Lalithabai EK, Mallika VK, Minimol JS, Abraham K. Cocoa in India. Thrissur: Director of Extension, Kerala Agricultural University; 2009.

13. Tetteh DA, Obeng D. Cocoa pods and seeds characteristics of different hybrid varieties in Southeastern Ghana. Ecol Evol Biol. 2021:6(3):88-91. DOI: 10.11648/j.eeb.20210603.13.

14. Ajmal PM, Suma B, Minimol JS, Sujith SS. Genetic Variability Estimate of Selected Criollo X Forastero Cocoa (*Theobroma cacao* L.) Hybrids. Adv Life Sci. 2016:5(518):7400-02.

15. Deepa R, Balakrishnan S, Jegadeeswari V, Thiribhuvanamala G, Kumaravadivel N. Study of variability, heritability and genetic advance for yield attributing characters in cocoa clones. Pharma Innovation. 2019:8(5):466-69.

16. Ofori A, Padi FK, Ansah FO, Akpertey A, Anim-Kwapon GJ. Genetic variation for vigour and yield of cocoa (*Theobroma cacao* L.) clones in Ghana. Sci Hortic. 2016:213:287-93. https://doi.org/10.1016/j.scienta.2016.11.003

17. Evans JD. Straightforward statistics for the behavioural sciences. Thomson Brooks/Cole Publishing Co.;1996.

18. Adewale BD. Diversity, genetic stability and reproductive biology of African yam bean (*Sphenostylis stenocarpa* (Hochst. ex A. Rich.) Harms). PhD Thesis. Abeokuta, Nigeria: Federal University of Agriculture; 2012.

19. Veeresh S, Minimol JS, Suma B, Panchami P, Shilpa KS. Multivariate analysis of genetic diversity in Phytophthora pod rot resistant exotic cocoa germplasm. J Plant Crops. 2020:48(3):202-12. https://doi.org/10.25081/jpc.2020.v48.i3.6628

20. Adenugaa OO, Ariyob OJ. Diversity analysis of cacao (*Theobroma cacao*) genotypes in Nigeria based on juvenile phenotypic plant traits. Int J Fruit Sci. 2020:20(S3):S1348-59. https://doi.org/10.1080/15538362.2020.1786487

UNDER PEER REVIEW