Determining kinship pattern of robusta and arabica coffee clones using multivariate analysis

Rubiyo Rubiyo[1], Sudarsono Sudarsono[2], Muhammad Fuad Anshori[3*], Nicho Nurdebyandaru[4], Yovita Anggita Dewi[1], and Miftahur Rizqi Akbar[5]

[1]Indonesian Center for Agricultural Technology Assessment and Development, 16114, Bogor, West Java, Indonesia.
[2]IPB University, Faculty of Agriculture, Department of Agronomy and Horticulture, 16680, Bogor, Indonesia.
[3]Hasanuddin University, Agronomy Department, 90245, Makassar, Indonesia. *Corresponding author (fuad.anshori@unhas.ac.id).
[4]Indonesian Agency for Agricultural Research and Development, Indonesian Center for Agricultural Land Resources Research and Development, 16114, Bogor, Indonesia.
[5]Indonesian Coffee and Cocoa Research Institute, 68118, Jember, Indonesia.

ABSTRACT

Identification of kinship in various coffees in the Indonesian Industrial & Beverages Crops Research Institute (IIBCRI) is essential for the coffee breeding process in the future. The kinship analysis can be done by using multivariate analysis. This study aimed to identify the kinship pattern and specific morphological characteristics of the coffee clone collections by using multivariate analysis. The experiment was carried out through the observational method corrected with analysis of covariance. The plant materials used were five genotypes of robusta coffee (*Coffea canephora* Pierre ex A. Froehner) and three arabica coffee (*Coffea arabica* L.) genotypes. Robusta genotypes studied were BP 308, BP 436, BP 42, SA 237, and SA 203; and arabica genotypes were S 795, Kartika 1, and Kartika 2. The study used an IPGRI (International Plant Genetic Resources Institute) list descriptor modified for coffee plants characterization. The morphological characters consisted of 46 characters. The results showed that coffee clones in the IIBCRI had relatively high diversity in clone grouping, both between types and between clones within the species. The grouping results showed that subgroup 1 of the robusta coffee consisted of BP 308, SA237, and SA203 coffee clones. Conversely, subgroups 2 of the robusta coffee consisted of BP 436 and BP 42. As for arabica coffee, subgroup 1 consisted of Kartika 1 and Kartika 2 clones, while the S795 clone has grouped solely in subgroup 2. Based on this study, the IIBCRI was also considered suitable for selecting and identifying morphological lines of coffee, particularly robusta coffee.

Key words: *Coffea* spp., dendrogram analysis, principal component, principal coordinates.

INTRODUCTION

Coffee (*Coffea* spp.) is one of the plantation crops developed in Indonesia since the Netherlands' colonialism. This crop has become a valued commodity in the strengthening of foreign exchange. It is represented by data of production, export, and coffee plantation area in Indonesia (Andini et al., 2021). Indonesia's coffee production has ranked the fourth rank below Brazil, Vietnam, and Colombia. Indonesia exports less than 0.28 million tons of coffee beans. The plantation area reaches 1.19 million hectares. The plantations are dominated by 96% smallholders and the remains from private and national corporations (Statistics Indonesia, 2019).

The kinds of coffee developed in Indonesia are arabica (*Coffea arabica* L.) and robusta (*Coffea canephora* Pierre ex A. Froehner). Both coffees have a demand level higher than the other kinds (Statistics Indonesia, 2019). Arabica coffee has the best taste among others (Dias and Benassi, 2015). The arabica plantation areas are less extensive than robusta coffee, but its economic values are more expensive than robusta coffee (Dias and Benassi, 2015; Nugroho et al., 2016). In general, arabica coffee is more planted at the highland around 800-2000 m a.s.l. (Nugroho et al., 2016; Konieczka et al., 2020). The taste and

resistance to *Hemileia vastatrix* attack are the reasons why arabica is grown in the highland (Alnopri and Hermawan, 2015). Robusta coffee is more resistant to *H. vastatrix*, it can be planted at less 1000 or 200-900 m a.s.l. and optimal at 600-700 m a.s.l. (Konieczka et al., 2020). The difference of agroecosystems between each coffee impacts different morphology characters and indirectly on the taste of both coffees. The coffee taste is mainly influenced by genetic interaction and agroecosystem (Bicho et al., 2013; Ramadiana et al., 2018), some coffees have a certificate as specialty coffee (Traore et al., 2018). Based on their unique characters, specific identification is interesting to differentiate between arabica and robusta in the same environment.

Indonesian Industrial & Beverages Crops Research Institute (IIBCRI) is a research institution focused on developing industrial crops, including coffee. This institution is located at Pasir Kuda, Sukabumi District, West Java, with an altitude of less than 400 m a.s.l. Based on the altitude, this environment is not optimal for the coffee plant; however, this institution has some robusta and arabica coffee collections in the same ecosystem. Both coffees have good growth at IIBCRI, and coffee collections do not present leaf rust diseases (Anshori, 2014). It is interesting to research genetic distance patterns or the differential response between coffees in the IIBCRI plantation environment. One of the simple methods to identify the difference of response between both coffees is morphological phenotype analysis. This method was reported by Ramadiana et al. (2018) for 15 robusta coffee accessions at Lampung, Indonesia. However, accuracy and a systematic approach are needed to minimize data bias in the analysis process, especially to many morphology characters (Mattjik and Sumertajaya, 2011). One approach that can be done is multivariate analysis.

Multivariate analysis is an approach to compress and combine big variable series, simple and easy to understand. This approach analysis consists of specific analyses categorized based on its function to reduce variance dimension to the cluster or to know the dependent relationship among all variables (Mattjik and Sumertajaya, 2011; Mengual-Macenlle et al., 2015). The standard analysis applied in the genetic distance is cluster analysis concepts. The kind of cluster analysis concepts that could be used is cluster gram analysis, principal component analysis (PCA), multiple dimension scaling, and principal coordinate analysis (PCoA) (Evgenidis et al., 2011; Tounekti et al., 2017; Akpertey et al., 2019). Therefore, applying some clustering analyses in this study would be a solution to detect genetic distance among coffee clones at IFCRI. This study aimed to identify the kinship pattern and specific morphological characteristics of the coffee clone collections by using multivariate analysis.

MATERIALS AND METHODS

The experiment was carried out at Indonesian Industrial & Beverages Crops Research Institute (IIBCRI), Sukabumi, West Java, Indonesia, from March until September 2014. In general, the IIBCRI field has an area of approximately 159.6 ha with flat to undulating topography. The soil types is latosol with a pH ranging from 5-6. The location has a type B climate according to Schmidt-Ferguson. The specific climate information on the study location is shown in Table 1. Plant materials consisted of five clones of robusta coffee (*Coffea canephora* Pierre ex A. Froehner), BP 308, BP 436, BP 42, SA 237 and SA 237, and three clones of arabica coffee (*Coffea arabica* L.), S 795, Kartika 1 and Kartika 2. All clones derived from Indonesian Coffee and Cocoa Research Institute, Jember, Indonesia. The experiment was conducted through observation, and all quantitative data were justified through covariance analysis according to Anshori (2014).

The observation used the descriptor list of IPGRI (1996) modified for the coffee plant. Forty-six morphology descriptors were evaluated for vegetative, inflorescence and flower, fruit, and seed. Vegetative description covered (17): overall appearance, plant habit, vegetative development, branch angle, leaf length, leaf width, petiole length, leaf shape, leaf apex shape, stipule length, internode, leaf petiole color, young leaf color, mature leaf color, stipule shape, firmness of leaf surface waves, and firmness of leaf edge waves; inflorescence and flower description covered (15): inflorescence position, inflorescence on old wood, anther insertion, flower tip shape, flower base shape, number of flowers per axil, number of fascicles per axil, number of flowers per fascicle, corolla tub length, number of stamens, corolla length, corolla width, anther length, stigma length, and pistil length; fruit description covered (7): fruit color, fruit shape, fruit-disc shape, endocarp texture, fruit length, fruit width, and pulp thickness; and seed description covered (7): husk seed length, bean length, husk seed width, bean width, husk seed thickness, bean thickness, and husk seed shape. The observation used the common observation tools, like meter tool, ruler, vernier calipers, pen, and digital scales to identify the quantitative characters. Besides that, the additional tool used in this study was the color chart from the Royal Horticultural Society (London, UK) to observe color morphology characters.

Table 1. Climate information at Indonesian Industrial & Beverages Crops Research Institute (IIBCRI).

Month	Tmin	Tmax	Tavg	RH_avg	RR	ds
	°C			%	mm	h
March	20.88	29.09	23.58	87.00	287.70	4.12
April	20.97	30.27	24.19	84.74	514.50	5.74
May	21.06	30.23	24.28	85.06	296.40	5.76
June	21.32	30.03	24.47	82.65	84.90	5.35
July	20.56	29.32	23.78	82.84	351.30	5.82
August	19.94	29.93	23.72	80.28	539.80	5.79
September	19.35	31.39	24.32	72.61	21.80	7.56
Average	20.58	30.04	24.05	82.17	299.49	5.74

Tmin: Minimum temperature; Tmax: maximum temperature; Tavg: average temperature; RH: relative humidity; RR: rainfall; ds: duration of the sunlight.

Data analysis

The data obtained from the observations were divided into qualitative data (22 characters) and quantitative data (24 characters). Quantitative data were selected first with principal component analysis (PCA) and Biplot PCA analysis. Character groups that had the same direction in the biplot analysis were chosen by looking at the slices of the dominant variance character on PC1 and PC2. Both analyses were performed with STAR 2.0.1 software (International Rice Research Institute [IRRI], Los Baños, Laguna, Philippines). After the quantitative characters were selected, all qualitative characters were combined with the selected quantitative data to analyze the kinship between IIBCRI coffee clones. The kinship analysis was carried out through cluster analysis and principal coordinate analysis (PCoA) with Rstudio software (RStudio, Boston, Massachusetts, USA). Cluster analysis used package cluster and PCoA analysis using ape and vegan packages. After that, the results of the PCoA biplot analysis were carried out by orthogonal mapping according to the win-win-where concept to determine the specific character of the clustered clones using paints software.

RESULTS

Identification of variance determinants based on PCA of quantitative characters

The results of the PCA analysis showed that the two PCs were able to represent the diversity of existing quantitative data; it could be seen from the variance accumulation, which reached 84% (Table 2). Based on PC1, fruit length and pulp thickness characters had a positive vector direction and were different from most of the other quantitative characters. On the other hand, the essential negative characters in PC1 were leaf length, leaf width, internodes, number of flowers per axil, number of flowers per fascicle, number of fascicles per axil, anther length, corolla length, pistil length, corolla tube length, corolla width, bean length, bean width, bean thickness, husk seed width, and husk seed thickness. Based on PC2, stipule length, number of flowers per fascicle, bean thickness, and husk seeds width were the characters with the most significant negative variety. On the other hand, the essential positive characters of PC2 were petiole length, stigma length, corolla width, number of stamens, fruit length, pulp thickness, bean length, and husk seed length.

Kinship analysis based on cluster analysis of eight coffee clones at the IIBCRI

The qualitative characters also were filtered by variance pattern among all clones. Based on the phenotype of qualitative characters, almost all characters had variance among all clones, except plant habit, leaf petiole color, mature leaf color, inflorescence position, anther insertion, and fruit-disc shape (Table 3). Cluster analysis result showed that two large groups had a dissimilarity distance of 70% (Figure 1). The two groups could be identified as the two types of coffee groups tested at the IIBCRI. It explained that the cluster analysis could adequately separate the two types with a distant kinship value. The first group (BP 308, SA 237, SA 203, BP 436, and BP 42) was identified as robusta coffee clones, while the second group (S795, Karika 1, and Karika 2) was identified as arabica coffee clones. Based on the cluster analysis, each group could be divided into two sub-groups at a dissimilarity distance of 40% or 0.4.

Table 2. Principal component analysis of 24 quantitative characters of coffee clone collections in Indonesian Industrial & Beverages Crops Research Institute (IIBCRI).

Characters	PC1	PC2	PC3	PC4
Leaf length	-0.23	-0.03	0.03	-0.07
Leaf width	-0.23	0.02	0.03	-0.01
Petiole length	-0.13	0.22	0.51	-0.14
Stipule length	-0.19	-0.20	-0.37	0.03
Internode	-0.22	0.07	0.13	-0.22
Number of flowers per axil	-0.23	-0.06	0.00	0.04
Number of flowers per fascicle	-0.22	-0.14	-0.16	-0.08
Number of fascicles per axil	-0.22	-0.02	0.01	0.11
Anther length	-0.22	0.09	0.09	0.15
Corolla length	-0.22	-0.05	-0.09	0.14
Pistil length	-0.22	0.02	-0.04	0.07
Stigma length	-0.19	0.35	0.17	0.14
Corolla tub length	-0.18	-0.01	0.37	0.29
Corolla width	-0.20	0.13	0.11	0.36
Number of stamens	-0.09	0.17	-0.29	0.58
Fruit length	0.05	0.60	-0.27	-0.30
Fruit width	-0.20	0.04	0.12	-0.28
Pulp thickness	0.18	0.40	0.13	0.08
Bean length	-0.20	0.23	-0.22	-0.03
Bean width	-0.22	-0.06	-0.13	-0.09
Bean thickness	-0.22	-0.15	0.01	-0.12
Husk seed length	-0.17	0.30	-0.30	-0.13
Husk seed width	-0.22	-0.11	-0.04	-0.16
Husk seed thickness	-0.22	-0.05	0.09	-0.20
Standard deviation	4.37	1.37	1.32	1.17
Proportion of variance	0.76	0.08	0.07	0.05
Cumulative proportion	0.76	0.84	0.91	0.96
Eigen values	19.07	1.88	1.74	1.37

PC: Principal component.

Kinship analysis based on PCoA of eight coffee clones at the IIBCRI, Sukabumi

The PCoA biplot analysis results showed that all clones were mapped according to the groups in the cluster analysis. However, in this analysis, the specific differences among coffee clones could be seen with character vectors, especially in robusta coffee (Figure 2). BP42 and BP436 as subgroup 2 in robusta coffee were in quadrants that differed from other robusta clones. On the other hand, all clones were in the same quadrant in arabica coffee, although the S795 clones have a clear distance from Kartika 1 and Kartika 2 if the mapping was based on the concept of win-win where it could be explicitly identified the supporting characters of each clone sub-group. Based on the win-win where concept, the arabica coffee sub-group consisted of Kartika 1 and Kartika 2 clones were grouped specifically by leaf apex shape (LAS), flower tip shape (FTS), fruit color (FC), endocarp texture (ET), fruit length (FL), pulp thickness (PT), bean length (BL), and husk seed shape (HSS), while the S795 clone was grouped specifically for the firmness of leaf edge waves (FLEW) and fruit shape (FS). As for robusta coffee, BP 308, SA 237, and SA 203 clones also have some specific characteristics such as the firmness of the leaf surface wave (FLSW), stipule shape (SS), and the flower base shape (FBS). On the other hand, the clones BP 436 and BP 42 were grouped specifically by young leaf color (YLC), number of flowers per fascicle (NFF), inflorescence on old wood (IOW), leaf shape (LS), vegetative development (VD), bean thickness (BT), overall appearance (OA), corolla width (CW), branch angle (BA) and husk seed width (HSW). The concept of win-win is a mapping concept based on connecting lines between the farthest clones, and then the lines are divided based on their orthogonal properties from the center point.

Table 3. Morphology phenotypes of all coffee clones in the Indonesian Industrial & Beverages Crops Research Institute (IIBCRI).

Nr	Characters	Robusta coffee					Arabica coffee		
		BP 308	BP 436	BP 42	SA 237	SA 203	S 795	Kartika 1	Kartika 2
		Vegetative							
1	Overall appearance	Pyramidal	Pyramidal	Pyramidal	Pyramidal	Pyramidal	Elongated conical	Elongated conical	Elongated conical
2	Plant habit	Shrub	Shrub	Shrub	Shrub	Shrub	Shrub	Shrub	Shrub
3	Vegetative development	Sympodial	Sympodial	Sympodial	Sympodial	Sympodial	Monopodial	Monopodial	Monopodial
4	Branch angle	Semi-erect	Semi-erect	Semi-erect	Semi-erect	Semi-erect	Horizontal	Semi-erect	Semi-erect
5	Leaf shape	Ovate	Elliptic	Elliptic	Ovate	Ovate	Ovate	Ovate	Elliptic
6	Leaf apex shape	Apiculate	Spatulate	Spatulate	Apiculate	Apiculate	Apiculate	Apiculate	Spatulate
7	Leaf petiole color	Green	Green	Green	Green	Green	Green	Green	Green
8	Young leaf color	Brownish	Brownish	Brownish	Brownish	Brownish	Green	Greenish	Greenish
9	Mature leaf color	Green	Green	Green	Green	Green	Green	Green	Green
10	Stipule shape	Triangular	Ovate	Ovate	Triangular	Triangular	Ovate	Ovate	Ovate
11	Firmness of leaf surface waves	Present	Absent	Absent	Present	Present	Present	Absent	Absent
12	Firmness of leaf edge waves	Present	Absent	Present	Present	Present	Present	Present	Present
13	Leaf length, cm	26.30	26.41	26.19	26.66	24.76	15.63	15.34	15.36
14	Leaf width, cm	11.11	11.09	11.58	11.30	11.12	5.64	5.91	6.32
15	Petiole length, cm	1.29	1.70	1.83	1.29	1.40	1.21	1.29	1.22
16	Stipule length, cm	0.89	0.52	0.51	0.78	0.81	0.38	0.34	0.33
17	Internode, cm	7.51	7.61	8.09	6.53	6.29	3.86	3.72	3.87
		Inflorescence							
18	Inflorescence position	Axillary	Axillary	Axillary	Axillary	Axillary	Axillary	Axillary	Axillary
19	Inflorescence on old wood	Present	Present	Present	Present	Present	Absent	Absent	Absent
20	Anther insertion	Excluded	Excluded	Excluded	Excluded	Excluded	Excluded	Excluded	Excluded
21	Flower tip shape	Acute	Acute	Acute	Acute	Acute	Acuminate	Acuminate	Acuminate
22	Flower base shape	Roll	Not roll	Not roll	Roll	Roll	Not roll	Not roll	Not roll
23	Number of flowers per axil	50.00	49.90	48.20	52.30	51.50	20.90	23.80	21.90
24	Number of flowers per fascicle	6.30	5.50	5.30	6.30	5.80	3.80	3.10	3.30
25	Number of fascicles per axil	9.70	9.60	9.40	9.80	10.20	5.40	6.40	6.50
26	Anther length, cm	1.04	1.13	1.18	1.20	1.11	0.73	0.85	0.83
27	Corolla length, cm	1.89	1.81	1.70	2.15	1.83	1.05	1.26	1.21
28	Stigma length, cm	3.24	3.05	2.93	3.26	2.97	1.77	2.22	2.21
29	Pistil length, cm	0.72	0.79	0.90	0.80	0.72	0.38	0.57	0.69
30	Corolla tub length, cm	1.06	1.34	1.13	1.18	1.18	0.80	1.03	1.01
31	Corolla width, cm	0.51	0.55	0.55	0.59	0.56	0.41	0.48	0.49
32	Number of stamens	5.10	5.00	5.20	5.20	5.30	5.10	5.20	5.10
		Fruit							
33	Fruit shape	Obovate	Roundish	Ovate	Obovate	Obovate	Ovate	Obovate	Elliptic
34	Fruit-disc shape	Not marked	Not marked	Not marked	Not marked	Not marked	Not marked	Not marked	Not marked
35	Fruit color	Orange-red	Red	Red	Orange-red	Red	Orange-red	Red	Red
36	Endocarp texture	Coriaceous	Coriaceous	Coriaceous	Coriaceous	Coriaceous	Coriaceous	Subcoriaceous	Subcoriaceous
37	Fruit length, cm	1.62	1.48	1.68	1.57	1.51	1.59	1.57	1.66
38	Fruit width, cm	1.52	1.51	1.53	1.46	1.43	1.36	1.39	1.31
39	Pulp thickness, cm	0.08	0.09	0.12	0.08	0.07	0.11	0.16	0.18
		Seed							
40	Husk seed shape	Elliptic	Ovate	Elliptic	Elliptic	Ovate	Ovate	Elliptic	Elliptic
41	Husk seed length, cm	1.32	1.20	1.30	1.32	1.20	1.15	1.19	1.19
42	Husk seed width, cm	0.97	0.95	0.93	0.95	0.93	0.83	0.79	0.81
43	Husk seed thickness, cm	0.62	0.64	0.62	0.62	0.57	0.49	0.46	0.46
44	Bean length, cm	1.08	0.95	1.06	1.05	1.00	0.84	0.90	0.91
45	Bean width, cm	0.81	0.77	0.80	0.82	0.79	0.69	0.66	0.65
46	Bean thickness, cm	0.51	0.51	0.48	0.50	0.49	0.39	0.36	0.36

Figure 1. Cluster analysis of coffee clone collections in the Indonesian Industrial & Beverages Crops Research Institute (IIBCRI)
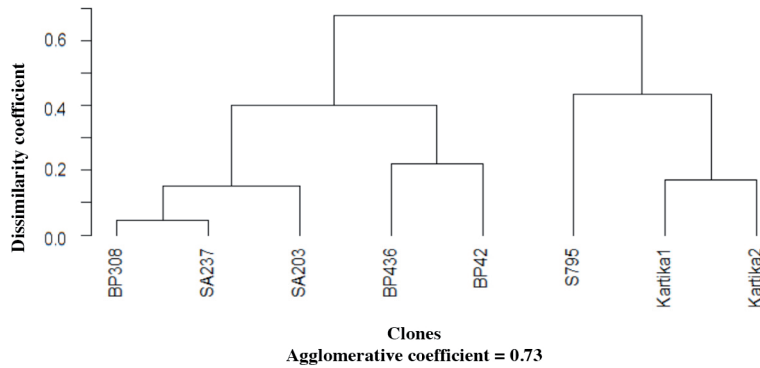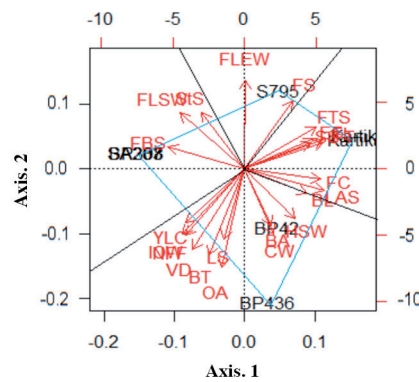

Figure 2. Principal coordinate analysis (PCoA) biplot analysis of coffee clone collections in the Indonesian Industrial & Beverages Crops Research Institute (IIBCRI)

LAS: Leaf apex shape; FTS: flower tip shape; FC: fruit color; ET: endocarp texture; FL: fruit length; PT: pulp thickness; BL: bean length; HSS: husk seed shape; FLEW: the firmness of leaf edge waves; FS: fruit shape; FLSW: the leaf surface wave; SS: stipule shape; FBS: the flower base shape; YLC: young leaf color; NFF: number of flowers per fascicle; IOW: inflorescence on old wood; LS: leaf shape; VD: vegetative development; BT: bean thickness; OA: overall appearance; CW: corolla width; BA: branch angle; HSW: husk seed width.

DISCUSSION

In general, PCA is an analysis that can reduce a large character dimension to be simpler by retaining most of the diversity information in the initial data (Mattjik and Sumertajaya, 2011; Jolliffe and Cadima, 2016; Wang et al., 2021). It was very efficient to see the dominant character proportionally affecting the population of the observed genotype. The use of this analysis has been reported by Anshori et al. (2021) and Farid et al. (2021) in determining essential characters in a population. Identifying important characters in PCA was carried out by looking at the dominant characters on the two most significant PCs independently (Jolliffe and Cadima, 2016; Tounekti et al., 2017). Based on this study, the cumulative proportion of two PCs value was more significant than the minimum limit stated by Mattjik and Sumertajaya (2011) of 80%. Thus, these PCs could be used in looking the dominant characters. Eigenvector of the most significant PCA loading, both positive and negative, is the basis for determining the dominant character (Mattjik and Sumertajaya, 2011; Tounekti et al., 2017; Anshori et al., 2021; Farid et al., 2021), then these results are combined to determine the main characteristic character.

The determination of selected characters was based on the eigenvector threshold for each PC. In PC1, the -0.2 value was used as the threshold. It is due to the eigenvector distribution spread on the negative quadrant, exactly on intervals 0 to -0.3, although one outlier has a more 0 value in PC1. So, the selected character was focused on less than -0.2. On the

other hand, in PC2, < -0.1 or > 0.1 were used as a threshold. The PC2 eigenvector distribution was relatively more varied than PC1. Interval of -0.1-0.1 was considered not more significant in determining PC2 variance. According to Mattjik and Sumertajaya (2011), a character more close to 0 indicated a meaningless of the character role, so this character was not considered in the subsequent step analysis. Therefore, a value > 0.1 or < -0.1 was used as the basis for determining the essential characters on PC2. Based on the second PC slices results, it was found that the number of flowers per fascicle, corolla width, fruit length, pulp thickness, bean length, bean thickness, and husk seed width were important quantitative characters with a large variety of directions. Therefore, these seven characters can be combined with qualitative characters in identifying kinship distances between coffee clones at the IIBCRI, Sukabumi.

Quantitative character selection was carried out to identify essential characters that influenced the diversity of clones observed. In general, identification of the genotype kinship is closely related to the nature of the observed character data. There are several data types on a statistical scale, namely binary, nominal, ordinal, and numeric (Ranganathan and Gogtay, 2019). Qualitative character or categorical data has a scope in discrete binary, nominal and ordinal data, although some numeric data also contains discrete numerical data (Ali and Bhaskar, 2016). The nature of discrete data will form a graph with a rigid gap (Ali and Bhaskar, 2016; Ranganathan and Gogtay, 2019) and is excellent in the process of differentiating a trait between genotypes. Genetically, discrete characteristics are influenced by only a few significant genes, so their characters are less affected by the dominance of the environment. The fewer groups of characters, the fewer genes that regulate the trait (Acquaah, 2012), and it is easier for qualitative characters to distinguish between genotypes.

In contrast, quantitative characters are continuous and controlled by many genes (Li et al., 2017), and the environment greatly influences the traits of a genotype unless most of the genes in these traits have the same variance response direction (Acquaah, 2012). Based on their genetic characteristics, coffee plants are relatively highly heterozygous because the propagation of coffee plants is mainly done by cloning of high hybrid vigor plants (Mohammed, 2011; Geneti, 2019) so that their characteristics will be more diverse between genotypes in an environment. Therefore, the quantitative characters in this study need to be filtered first to get the main characters with a large diversity.

The variable should have a variance pattern among the objective in kinship identification. It could help to know the specific characters which determine the object mapping, especially in the cluster and PCoA. So, based on this, plant habit, leaf petiole color, mature leaf color, inflorescence position, anther insertion, and fruit-disc shape were not included in cluster analysis and PCA.

In general, cluster analysis is one of the multivariate analyses used to visualize the relationships between objects based on the similarity or dissimilarity distances of various characters (Mattjik and Sumertajaya, 2011). Several studies widely use this analysis to determine the genetic distance between each genotype of a population, using both quantitative characters (Tounekti et al., 2017; Malau and Pandiangan, 2018; Anshori et al., 2020), qualitative characters (Henry et al., 2015), metabolomic data (Hanifah et al., 2018), molecular data (Li et al., 2017) and the combination of each character (Tessema et al., 2011; Solankey et al., 2015; Kachare et al., 2020). Based on the cluster analysis results, the coffee clones found in the IFCRI germplasm collection have relatively high diversity, both in arabica and robusta coffee. Clones with close kinship are BP 308 and SA 237 in the robusta coffee type, so crossing between these coffees is not recommended because they have close kinship distances.

The PCoA was performed to refine and detail the clustering performed by cluster analysis. Cluster analysis can only identify kinship distances. However, cluster analysis cannot explain how the kinship is formed from many variables (Evgenidis et al., 2011; Anshori et al., 2020). PCA and PCoA cover this deficiency to understand the formed groupings comprehensively. Several studies have also reported this concept in complementing the information contained in cluster analysis (Evgenidis et al., 2011; Tounekti et al., 2017). The PCoA biplot analysis is preferred in this study over the PCA biplot analysis. In general, the difference between PCA and PCoA lies in the basis of the matrix in making eigenvalues and eigenvectors. The PCA is based more on correlation or covariance analysis, while PCoA is based on kinship value matrices (Mattjik and Sumertajaya, 2011). The PCoA allows various categorical data, especially those with many 0 values, or a mixture of categorical and parametric can be distributed linearly like PCA analysis (Liu et al., 2019). The results of PCA analysis are in line with the data on dendrogram formation, mostly categorical data. Therefore, kinship values are better used in the root characteristic formation than correlation or covariance, which is more identical to quantitative data.

Based on cluster analysis and PCoA, the diversity of characters in the robusta coffee type is higher than the diversity in arabica coffee. The robusta coffee is a cross-pollinating (Ramadiana et al., 2018), while the arabica is a self-pollinating plant (Déchamp et al., 2015). In addition, arabica coffee generally grows the best above 800 m a.s.l. (Nugroho et al., 2016; Konieczka et al., 2020). Cross pollinations make robusta coffee more diverse than arabica coffee. In addition, based on the results of kinship analysis, the environment of the IFCRI was considered quite good in mapping the traits between coffee clones so that the environment could be recommended as an identification environment for the diversity of coffee accessions. This study results need further evaluation for the taste-associated characters and a larger population to increase the result's precision.

CONCLUSIONS

At the Indonesian Industry and Freshner Crops Research Institute (IFCRI), Sukabumi, coffee clones show high diversity in clone grouping, both between types and between clones within types. The clustering results showed that Subgroup 1 of the first robusta coffee consisted of BP 308, SA237, and SA203 coffee clones with specific characters in the firmness of the leaf surface wave, stipule shape, and flower base shape. Subgroup 2 of robusta coffee consisted of BP 436 and BP 42 clones with specific characters in young leaf color, flower number per fascicle, inflorescence on old wood, leaf shape, vegetative development, bean thickness, and overall bean appearance, corolla width, branch angle, and husk seed width. Sub-group 1 on arabica coffee consisted of the S795 clone with the firmness of leaf edge waves and fruit shape as its specific characters. In contrast, the second sub-group of arabica coffee consisted of Karika 1 and Karika 2 clones with particular characters: leaf apex shape, flower tip shape, fruit color, endocarp texture, and fruit length, pulp thickness, bean length, and husk seed shape. Based on this research, it can also be concluded that the environment of the IFCRI is considered suitable for the selection and identification of the morphological lines of coffee, especially robusta coffee.

ACKNOWLEDGEMENTS

REFERENCES

Acquaah, G. 2012. Principles of plant genetics and breeding. 2nd ed. Blackwell Publishing, Oxford, UK.

Akpertey, A., Anim-Kwapong, E., and Ofori, A. 2019. Assessment of genetic diversity in robusta coffee using morphological characters. International Journal of Fruit Science 19(3):276-299. doi:10.1080/15538362.2018.1502723.

Ali, Z., and Bhaskar, S.B. 2016. Basic statistical tools in research and data analysis. Indian Journal of Anaesthesia 60:662-9.

Alnopri, and Hermawan, B. 2015. Sustainability for growth and productivity of arabica coffee in lowland regions of Bengkulu Province. International Journal on Advanced Science Engineering Information Technology 5(5):304-307.

Andini, R., Muzaifa, M., Marlina, L., Sulaiman, M.I., Jaya, R., Muslih, A.M., et al. 2021. Making biodiversity work for coffee production. A case study of Gayo arabica coffee in Indonesia. MOJ Ecology & Environmental Sciences 6(4):156-162.

Anshori, M. 2014. Analisis Keragaman morfologi koleksi tanaman kopi arabika dan robusta Balai Penelitian Tanaman Industri dan Penyegar Sukabumi. IPB University, Bogor, Indonesia. [Indonesian]

Anshori, M.F., Bambang, S.P., Iswari, S.D., Sintho, W.A., and Willy, B.S. 2021. A new approach to select doubled haploid rice lines under salinity stress using indirect selection index. Rice Science 28:368-378. doi:10.1016/j.rsci.2021.05.007.

Anshori, M.F., Purwoko, B.S., Dewi, I.S., Ardie, S.W., and Suwarno, W.B. 2020. Cluster heatmap for detection of good tolerance trait on doubled-haploid rice lines under hydroponic salinity screening. IOP Conference Series: Earth and Environmental Science 484:012001.

Bicho, N.C., Lidon, F.C., Ramalho, J.C., and Leitão, A.E. 2013. Quality assessment of arabica and robusta green and roasted coffees – A review. Emirates Journal of Food Agriculture 25(12):945-950.

Déchamp, E., Breitler, J.C., Leroy, T., and Etienne, H. 2015. Coffee (*Coffea arabica* L.) p. 275-291. In Wang, K. (ed.) *Agrobacterium* protocols. Vol. 2. Methods in molecular biology 1224. Springer Science+Business Media, New York, USA. doi:10.1007/978-1-4939-1658-0_22.

Dias, R.C.E., and Benassi, Md.T. 2015. Discrimination between arabica and robusta coffees using hydrosoluble compounds: is the efficiency of the parameters dependent on the roast degree? Beverages 1:127-139. doi:10.3390/beverages1030127.

Evgenidis, G., Traka-Mavrona, E., and Koutsika-Sotiriou, M. 2011. Principal component and cluster analysis as a tool in the assessment of tomato hybrids and cultivars. International Journal of Agronomy 2011:697879. doi:10.1155/2011/697879.

Farid, M., Nasaruddin, Anshori, M.F., Musa, Y., Iswoyo, H., and Sakinah, A.I. 2021. Interaction of rice salinity screening in germination and seedling phase through selection index based on principal components. Chilean Journal of Agricultural Research 81:368-377. doi:10.4067/S0718-58392021000300368.

Geneti, D. 2019. Review on heterosis and combining ability study for yield and morphological characters of coffee (*Coffea arabica* L.) in Ethiopia. Journal of Environment and Earth Science 9(12):24-29.

Hanifah, A., Maharijaya, A., Putri, S.P., Laviña, W.A., and Sobir. 2018. Untargeted metabolomics analysis of eggplant (*Solanum melongena* L.) fruit and its correlation to fruit morphologies. Metabolites 8(3):49. doi:10.3390/metabo8030049.

Henry, D., Dymnicki, A.B., Mohatt, N., Allen, J., and Kelly, J.G. 2015. Clustering methods with qualitative data: A mixed-methods approach for prevention research with small samples. Prevention Science 16(7):1007-1016. doi:10.1007/s11121-015-0561-z.

IPGRI. 1996. Descriptors for coffee (*Coffea* spp. and *Psilanthus* spp.) International Plant Genetic Resources Institute (IPGRI), Rome, Italy.

Jolliffe, I.T., and Cadima, J. 2016. Principal component analysis: a review and recent developments. Philosophical Transaction of the Royal Society A 374:20150202. doi:10.1098/rsta.2015.0202.

Kachare, S., Tiwari, S., Tripathi, N., and Thakur, V.V. 2020. Assessment of genetic diversity of soybean (*Glycine max*) genotypes using qualitative traits and microsatellite markers. Agricultural Research 9(1):23-34. doi:10.1007/s40003-019-00412-y.

Konieczka, P.P., Aliaño-González, M.J., Ferreiro-González, M., Barbero, G.F., and Palm, M. 2020. Characterization of arabica and robusta coffees by ion mobility sum spectrum. Sensor 20:3123. doi:10.3390/s20113123.

Li, Y., Yang, K., Yang, W., Chu, L., Chen, C., Zhao, B., et al. 2017. Identification of QTL and qualitative trait loci for agronomic traits using SNP markers in the adzuki bean. Frontiers in Plant Science 8:840. doi:10.3389/fpls.2017.00840.

Liu, H., Hu, Z., Zhou, M., Hu, J., Yao, X., Zhang, H., et al. 2019. The distribution variance of airborne microorganisms in urban and rural environments. Environmental Pollution 247:898-906.

Malau, S., and Pandiangan, S. 2018. Morphological variation in arabica coffee (*Coffea arabica* L.) growing in North Sumatra Indonesia. Indonesian Journal of Agronomy 46(3):314-321.

Mattjik, A.A., and Sumertajaya, I.M. 2011. Multivariate analysis using SAS. Faculty of Mathematics and Natural Science, Bogor Agricultural University, Bogor, Indonesia.

Mengual-Macenlle, N., Marcos, P.J., Golpe, R., and González-Rivas, D. 2015. Multivariate analysis in thoracic research. Journal Thoracic Disease 7(3):E2-E6.

Mohammed, W. 2011. Heterosis and combining ability analysis for coffee quality in diallel crosses of diverse coffee (*Coffea arabica* L.) parents in origin. East African Journal of Sciences 5(1):12-21.

Nugroho, D., Basunanda P., and Mw, S. 2016. Physical bean quality of arabica coffee (*Coffea arabica*) at high and medium altitudes. Pelita Perkebunan 33:1-11.

Ramadiana, S., Hapsoro, D., and Yusnita, Y. 2018. Morphological variation among fifteen superior robusta coffee clones in Lampung Province, Indonesia. Biodiversitas 19(4):1475-1481.

Ranganathan, P., and Gogtay, N.J. 2019. An introduction to statistics - Data types, distributions and summarizing data. Indian Journal Critical Care Medicine 23:S169-S170. doi:10.5005/jp-journals-10071-23198.

Solankey, S.S., Singh, P.K., and Singh, R.K. 2015. Genetic diversity and interrelationship of qualitative and quantitative traits in sweet potato. International Journal of Vegetable Science 21(3):236-248. doi:10.1080/19315260.2013.867295.

Statistics Indonesia. 2019. Indonesian coffee statistics 2018. Statistics Indonesia, Jakarta, Indonesia.

Tessema, A., Alamerew, S., Kufa, T., and Garedew, W. 2011. Genetic diversity analysis for quality attributes of some promising *Coffea arabica* germplasm collections in Southwestern Ethiopia. Journal of Biological Sciences 11:236-244.

Traore, T.M., Wilson, N.L.W., and Fields, D. 2018. What explains specialty coffee quality scores and prices: a case study from the cup of excellence program. Journal of Agricultural and Applied Economics 50(3):349-368. doi:10.1017/aae.2018.5.

Tounekti, T., Mahdhi, M., Al-Turki, T.A., and Khemira, H. 2017. Genetic diversity analysis of coffee (*Coffea arabica* L.) germplasm accessions growing in the Southwestern Saudi Arabia using quantitative traits. Natural Resources 8:321-336.

Wang, R., Peng, W., and Wu, W. 2021. Principal component analysis and comprehensive evaluation on drought tolerance difference of canola cultivars at germination and emergence stages. Chilean Journal of Agricultural Research 81:557-567. doi:10.4067/S0718-58392021000400557.