Philippine Journal of Science
147 (4): 659-665, December 2018
ISSN 0031 - 7683
Date Received: 05 Apr 2018

Interspecific and Intraspecific Variation in *Coffea* sp. using Fruit Metric and Landmark-based Geomorphometric Analyses

Joana Camille S. Sanchez[1], **Jonas P. Quilang**[1], **Ruel M. Mojica**[2], and **Ernelea P. Cao**[1]*

[1]Institute of Biology, College of Science,
University of the Philippines, Diliman, Quezon City
[2]National Coffee Research Development and Extension Center,
Cavite State University, Indang, Cavite

Coffee is considered one of the most popular cash crops in the world. In the Philippines, three species of *Coffea* and correspondingly four varieties (*C. arabica* var. Red Bourbon, *C. canephora* var. Robusta, *C. liberica* var. Liberica, and *C. liberica* var. Dewevrei) are grown to produce coffee with each having a distinct taste and aroma. However, misidentification of farmers based on morphological traits is often a problem. In this study, an easy, fast, and inexpensive method for differentiating the four different varieties of *Coffea* was established using metric and geomorphometric analysis. At least 30 berries were collected for each variety from a single site at the Cavite State University in Indang, Cavite wherein metric characters were measured and analyzed using ANOVA, SIMPER, and ANOSIM. At the same time, geomorphometric analysis of the fruit shape – particularly the shape of the berry as oriented radially and longitudinally – was also done using CVAgen. For the metric analysis, the length of the minor radial axis was the most discriminating characteristic. In the geomorphometric analysis, the longitudinal view of *C. arabica* var. Red Bourbon, *C. canephora* var. Robusta, and *C. liberica* var. Liberica was found to differentiate them from each other. Contrastingly, *C. liberica* var. Dewevrei was the only variety that could not be distinguished in any geomorphometric analyses and can only be distinguished using metric analysis.

Key words: berry shape, coffee, geomorphometrics, major axis, metrics

INTRODUCTION

Coffea is a flowering plant under family *Rubiaceae,* subfamily *Ixoroideae* that comprises 103 species. The seeds from three *Coffea* species – namely, *C. arabica, C. canephora*, and *C. liberica* – are used in the production of coffee. They are usually found in humid evergreen forests, but some species can also be found in seasonally dry deciduous forests or bush lands (Maurin *et al*. 2007). The identification of *Coffea* is based on the presence of the following: (1) a single main trunk with plagiotrophic branching; (2) paired, axillary inflorescences; (3) truncate to undulate calyces; (4) hermaphrodite flowers; (5) white to lightly pink corollas; (6) exserted anthers; and (7) berries commonly with two seeds that have an invagination on their ventral side (Davis *et al*. 2006).

A number of problems have plagued the Philippine coffee industry. For one, the lack of proper identification of species and varieties being grown has resulted to farmers planting incorrect varieties and therefore losing out on selling certain varieties preferred by traders. Coffee is a popular beverage and a valuable export product (Anthony *et al*. 2001). It is the world's second most

*Corresponding author: epcao@up.edu.ph

traded commodity next to oil. Due to the high demand for coffee exports worldwide, farmers must however supply the correct varieties that are preferred by certain consumers. Arabica coffee is the most cultivated among the coffee species due to its superior taste quality, rich aroma, and low caffeine content (Mishra & Slater 2012), while Robusta coffee – with its more inferior taste – is resistant to many pests and diseases. Liberica coffee or *kapeng barako* has a very strong taste and flavor but is also resistant to many pests and diseases compared to Dewevrei or Excelsa coffee or *kapeng excelsa*, which has a better taste and aroma compared to Robusta and Liberica but is commercially cultivated in lesser quantities. In order to produce a desired variety for certain purposes, it is necessary to be able to maintain genetic consistency and identity for market purposes.

Morphological and molecular characterization allows discrimination and classification among the different varieties. However, identification using molecular techniques is expensive – as well as impractical – in the field or for commercial use. As such, morphological variation among the different varieties must be studied to establish techniques that can successfully differentiate one variety from another. Studying morphological variation will provide a method of rapid identification of species and hopefully even varieties. However, morphological characterization using qualitative descriptions *per se* may be subjective. On the other hand, quantitative data will provide more accurate characterization of the species and/or varieties for any organism. Metrics involve measurement of parts, while geomorphometrics determine variations in geometric shape. Geomorphometrics can be either outline-based or landmark-based. In outline-based geometric morphometrics, the edge boundaries of a sample are identified and analyzed using mathematical functions such as the Fourier analysis. The curves of these samples are then compared through their coefficients and visualized as outlines of their morphological shape. For landmark-based geometric morphometrics, landmarks that are biologically relevant are placed on the samples. These landmarks are assigned with coordinates. Landmark-based geometric morphometrics cannot be analyzed directly due to variations in size, position, and orientation that may obscure the true shape of the samples. As such, size, position, and orientation correction methods are used prior to analysis. Commonly, superimposition is done to compare the shape of organisms through the coordinates of their landmarks. Other methods for analysis are Euclidean distance matrix analysis (EDMA) and finite element scaling analysis (FESA) (Adams *et al*. 2004).

A number of studies have already utilized landmark-based geometric morphometric techniques to evaluate morphological variation. The study of Chong and co-workers (2014) used pollen grains to differentiate *C. arabica* var. Red Bourbon, *C. canephora* var. Robusta, *C. liberica* var. Liberica, and *C. liberica* var. Dewevrei, while the work of Cao and co-authors (2014) used coffee leaves to differentiate the same four varieties. A study that differentiates these four varieties based on fruit morphology, however, has not been done.

This work aimed to evaluate morphological variation among the fruits or berries of four varieties of three species of *Coffea* (*C. arabica* var. Red Bourbon, *C. canephora* var. Robusta, *C. liberica* var. Liberica, and *C. liberica* var. Dewevrei) using metric and landmark-based geomorphometric analyses and determine which character or part of the fruit can be utilized as a marker for identifying and differentiating each variety from one another.

MATERIALS AND METHODS

Sample Collection

Fruits, commonly known as berries, of four *Coffea* varieties – *C. arabica* var. Red Bourbon, *C. canephora* var. Robusta, *C. liberica* var. Liberica, and *C. liberica* var. Dewevrei – were collected from a single plantation site at the National Coffee Research Development and Extension Center (NCRDEC) of the Cavite State University in Indang, Cavite. Given that only one variety of *C. arabica* and *C. canephora* was obtained, *C. arabica* var. Red Bourbon – in this paper – will be referred to as simply *C. arabica,* while *C. canephora* var. Robusta will be called *C. canephora*. A minimum of 30 fruits for each variety was randomly selected throughout the course of the sampling. *Coffea* fruits were stored in an iced cooler during transfer to ensure freshness during data gathering. Samples were immediately processed *i.e*., measured and photographed upon reaching the Plant Genetics and Cyanobacterial Biotechnology Laboratory at the Institute of Biology, University of the Philippines in Diliman, Quezon City. The fruits were thereafter stored in a refrigerator set at 20 °C in the laboratory for future reference purposes.

Determination of Fruit Metric Characteristics and Measurements

Three characteristics were measured *i.e*., maximum longitudinal length, length of the major radial axis, and length of the minor radial axis (Figure 1). Ratios of these characteristics were also derived from the initial measurements.

Digitization of Samples for Fruit Geomorphometric Analysis

The *Coffea* fruits were photographed longitudinally and radially (Figures 2 and 3, respectively).

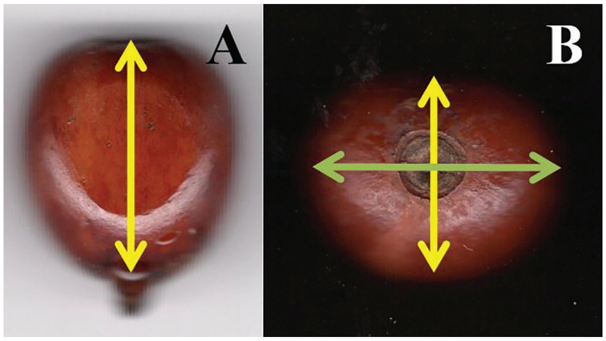

Figure 1. *Coffea liberica* var. Dewevrei fruits in different angles [A: longitudinal view (yellow line – maximum longitudinal length), B: radial view (green line – major radial axis, yellow line – minor radial axis)].

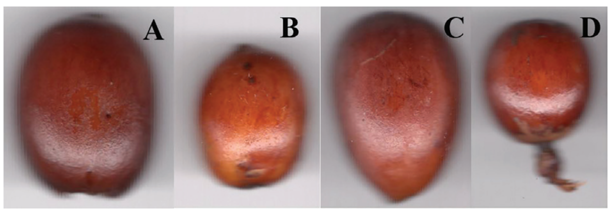

Figure 2. Representative photographs of the berries in longitudinal view of the four coffee varieties (A: Red Bourbon, B: Robusta, C: Liberica, D: Dewevrei).

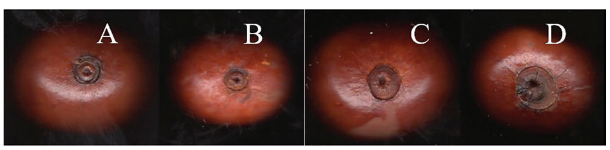

Figure 3. Representative photographs of the berries in radial view of the four coffee varieties (A: Red Bourbon, B: Robusta, C: Liberica, D: Dewevrei).

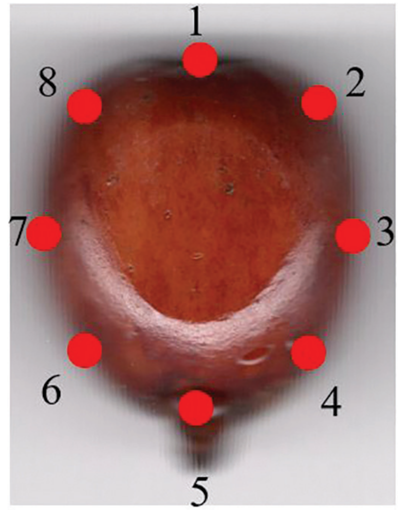

Figure 4. Longitudinally-oriented *Coffea liberica* var. Dewevrei fruit with landmark placement.

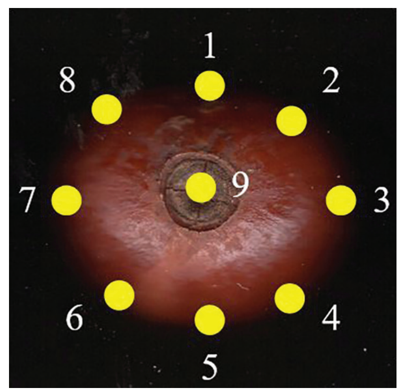

Figure 5. Radially-oriented *Coffea liberica* var. Dewevrei fruit with landmark placement.

Fruit specimens were digitized alongside a ruler as a reference to their size. Fruit samples were viewed longitudinally using a Perfection V39 flatbed scanner (Seiko Epson Corp., Nagano, Japan) with the resolution set at 300 dpi. Landmarks were placed, following the outline of the fruit specimens as shown in Figure 4. The first landmark was situated at the top proximal endpoint, while the other landmarks were placed clockwise around the specimen.

For radially-oriented fruit images, the disk or remnant of the fruit style was designated as the front face of the fruit specimens. These were digitized using a Perfection V39 flatbed scanner (Seiko Epson Corp., Nagano, Japan). Landmark placement was again made to follow the outline of the fruit specimens as demonstrated in Figure 5, with the last landmark – landmark 9 – being placed on the remnant of the fruit style.

Data Analysis

Once landmark placement was done, the landmarks of each image were superimposed on each other using Coordgen (H.D. Sheets, http://www2.canisius.edu/~sheets/morphsoft.html). Variations in size and orientation were corrected using the Procrustes method (Rohlf & Slice 1990). Briefly, in the Procrustes method, each configuration of landmarks is centered at the origin by subtracting the coordinates of its centroid from the coordinates of each landmark (Zelditch *et al*. 2012). The centroid is the center of the shape formed by the landmarks; its coordinates are obtained by getting the mean of all the X-coordinates and all the Y-coordinates. Afterwards, the landmark configurations are scaled to unit centroid size by dividing each coordinate of each landmark by the centroid size of that configuration. The centroid size is the square root of the sum of the squared distances of each landmark from the centroid. Lastly, one configuration is selected as reference then the second configuration is rotated to minimize the sum of the squared distances between homologous landmarks. This way, the effect of size differences between specimens and the effect of differences in orientation of images are removed.

Metric data were subjected to pairwise comparisons using the Kruskal-Wallis H Test and Mann-Whitney Post Hoc Test with SPSS 22 Software. The same data set was also subjected to further tests – Analysis of Similarity (ANOSIM) and Similarity Percentage (SIMPER) – using Past3 Software (Hammer *et al*. 2001). ANOSIM is a non-parametric test of significant difference between two or more groups based on distance measures that are converted to ranks (Clarke 1993). On the other hand, SIMPER is a method for determining which characters are primarily responsible for an observed difference between groups of samples (Clarke 1993). Lastly, geometric morphometric data were analyzed with Canonical Variate Analysis – Multivariate Analysis of Variance (CVA–MANOVA) using CVAgen (H.D. Sheets, http://www2.canisius.edu/~sheets/morphsoft.html). CVA is an ordination method that is used to simplify differences among groups and to form a discriminant function, which is used to assign specimens to groups (Zelditch *et al*. 2012). MANOVA is used to test significant differences in the mean forms of the groups using the CV scores generated from CVA. CVAgen was also used for validation *i.e*., to determine the *a posteriori* assignment of each specimen based on its distance from the group means and compare it with its *a priori* group assignment. PCAGen (H.D. Sheets, http://www2.canisius.edu/~sheets/morphsoft.html) was used to visualize deformation grid. A deformation grid is a graphical depiction of changes in shape. A deformation is a smooth function that maps landmarks in one form to corresponding landmarks in another form (Zelditch *et al*. 2012). A deformation grid points to which landmarks account for the differences in shapes between groups.

RESULTS

Morphological Measurements

Fruit Measurements. Table 1 shows the means of fruit measurements of the four coffee varieties. A significant difference in the means of the four varieties was obtained from Kruskal–Wallis H Test at a 95% confidence interval ($p<0.05$). Using Mann–Whitney U Test, the maximum longitudinal length of *C. arabica* and *C. liberica* var. Dewevrei was the only pairwise comparison that did not have a significant difference. Measurements on the radial side were found to be significant depending on the axis used. The length of the major axis of each coffee variety was significantly different from one another, while the minor axis was only significant if *C. arabica* or *C. canephora* were compared with *C. liberica* var. Liberica or *C. liberica* var. Dewevrei. Ratios MLL:MA1 and MLL:MA2 were found to be insignificant between *C. canephora* and *C. liberica* var. Dewevrei and also for *C. arabica* and *C. liberica* var. Liberica. The same trend could be seen for the MA1:MA2 ratio. Only *C. canephora* and *C. liberica* var. Liberica are also not significantly different from each other.

C. canephora and *C. liberica* var. Liberica were the most dissimilar in terms of fruit measurements. This value is followed closely by *C. canephora* and *C. liberica* var. Liberica. Dissimilarity values are shown in Table 2.

Table 1. Means of fruit measurements of *C. arabica*, *C. canephora*, *C. liberica* var. Liberica, and *C. liberica* var. Dewevrei.

Coffee Variety	Means					
	MLL	MA1	MA2	MLL: MA1	MLL: MA2	MA1: MA2
C. arabica var. Red Bourbon	1.5967[a]	1.2933[a]	1.0867[a]	1.2395[a]	1.4763[a]	1.1925[a]
C. canephora var. Robusta	1.2600[b]	1.1900[b]	1.0500[a]	1.0597[b]	1.1996[b]	1.1338[a,b]
C. liberica var. Liberica	1.9333[c]	1.5633[c]	1.3467[c]	1.2373[a]	1.4370[a]	1.1620[a]
C. liberica var. Dewevrei	1.6567[a]	1.5033[d]	1.3500[c]	1.1043[b]	1.2294[b]	1.1140[b]

*MLL – maximum longitudinal length, MA1 – major radial axis, MA2 – minor radial axis.
**In each column, means with the same superscript letter are not significantly different using Kruskal-Wallis H Test and Mann-Whitney U Test at a level of significance of 5%.

Table 2. Analysis of similarity (ANOSIM) of fruit measurements of *C. arabica*, *C. canephora*, *C. liberica* var. Liberica, and *C. liberica* var. Dewevrei.

Coffee Variety	*C. arabica* var. Red Bourbon	*C. canephora* var. Robusta	*C. liberica* var. Liberica	*C. liberica* var. Dewevrei
C. arabica var. Red Bourbon	-	0.6690	0.8550	0.6954
C. canephora var. Robusta	0.6690	-	0.9959	0.9010
C. liberica var. Liberica	0.8550	0.9959	-	0.8022
C. liberica var. Dewevrei	0.6954	0.9010	0.8022	-

The length of the radial minor axis of the coffee fruits was the factor that yielded the highest similarity between fruits from the same variety and greatest dissimilarity between group pairs (Table 3). This is followed closely by the ratio of the major longitudinal axis to the radial major axis. Together, they constitute approximately 70% of the total similarity of individuals in the same group. In contrast, the remaining two ratios – MLL:MA2 and MA1:MA2 – have the lowest contribution.

Table 3. Similarity percentages (SIMPER) of fruit measurements of *C. arabica*, *C. canephora*, *C. liberica* var. Liberica, and *C. liberica* var. Dewevrei.

	Contribution (%)	Cumulative Contribution (%)	Average Dissimilarity
MA2	36.530	36.530	6.035
MLL:MA1	34.280	70.810	5.663
MLL	11.980	82.780	1.979
MA1	8.027	90.810	1.326
MLL:MA2	6.534	97.350	1.079
MA1:MA2	2.655	100.000	0.439

Geomorphometrics

Longitudinal View of the Coffea Fruit. Geomorphometric analysis of the *Coffea* berry landmarks in longitudinal view yielded two distinct canonical variates, which had *p*-values <0.001. These indicate a significant difference in the shape of the *Coffea* berry when viewed longitudinally. All landmarks showed variation (Figure 6). In the CVA plot, 93% of *C. arabica*, 73% of *C. canephora*, 76% of *C. liberica* var. Liberica, and 50% of *C. liberica* var. Dewevrei were correctly classified (Figure 7). Most misclassified *C. liberica* var. Dewevrei were mistaken as *C. canephora* (36%).

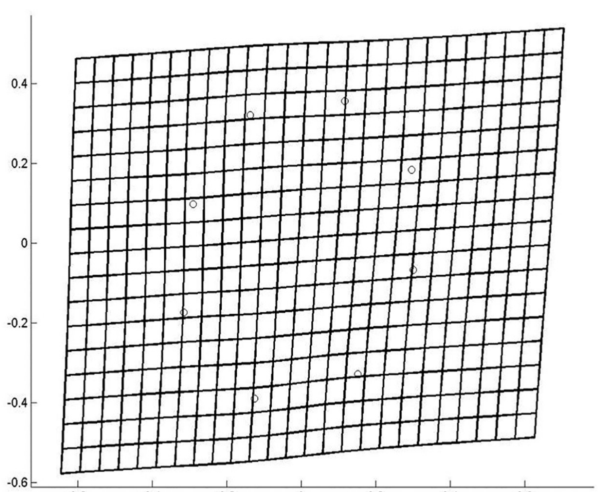

Figure 6. Deformation grid of *Coffea* fruit landmarks in longitudinal view.

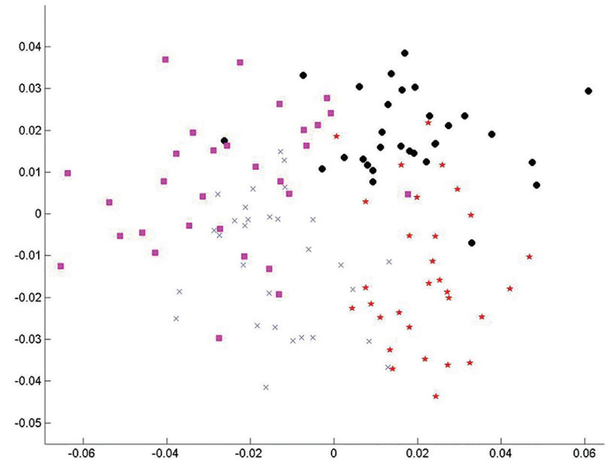

Figure 7. CVA plot of *Coffea* fruit landmarks in longitudinal view.

Radial View of the Coffea Fruit. Analysis of the shape of *Coffea* berries in radial view showed one distinct canonical variate (*p*<0.001). It was observed that there was a significant difference among varieties in terms of the shape of *Coffea* berries in radial view, given that the *p*-value obtained was less than 0.001. Variation was mostly due to differences in the radial major axis and in the position of the disk remnant, as the deformations in the grids are located at or near these landmarks (Figure 8). When grouped using CVA, 70% of *C. arabica*, 33% of *C. canephora*, 30% of *C. liberica* var. Liberica, and 56% of *C. liberica* var. Dewevrei were correctly classified (Figure 9). *C. canephora* was mostly misclassified as *C. liberica* var. Dewevrei (43%), while *C. liberica* var. Dewevrei was likewise mostly misclassified as *C. canephora* (23%). Many misclassifications of *C. liberica* var. Liberica incorrectly clustered with *C. arabica* (46%).

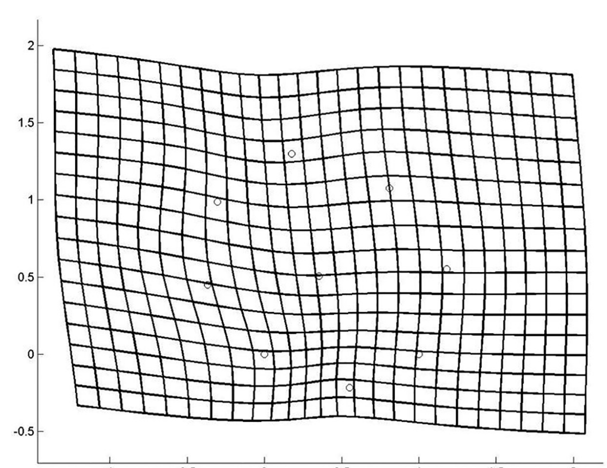

Figure 8. Deformation grid of *Coffea* fruit landmarks in radial view.

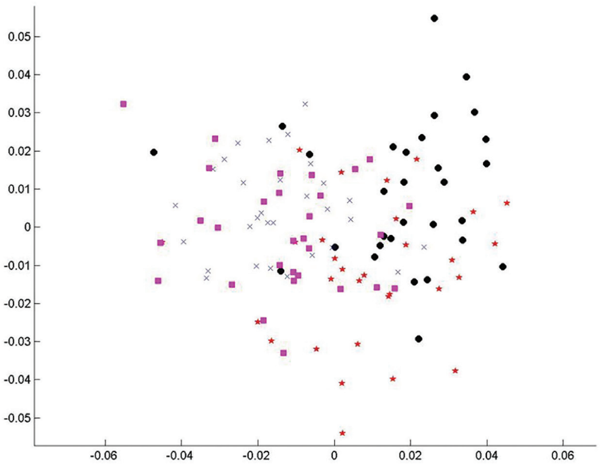

Figure 9. CVA plot of *Coffea* fruit landmarks in radial view.

DISCUSSION

Metric Data

The multivariate tests used – ANOVA, ANOSIM, and SIMPER – indicate that the fruit morphology of the four *Coffea* varieties are significantly different and can be distinguished from each other. All four *Coffea* varieties were found to be dissimilar from each other (R>0.5). The main contributing factor to this dissimilarity was the length of the minor radial axis, followed closely by the ratio of maximum longitudinal length to the length of major radial axis. Together, these make up 70.81% of the total dissimilarity between *Coffea* fruits. The length of the minor radial axis can be used to differentiate *C. arabica* and *C. canephora* from the two varieties of *C. liberica*, with the *C. liberica* varieties having a significantly longer minor radial axis. *C. arabica* from *C. canephora* and *C. liberica* var. Liberica from *C. liberica* var. Dewevrei can be further differentiated with the ratio of their maximum longitudinal length to the length of their major radial axis. *C. arabica* was more ellipsoid with a longer maximum longitudinal length than *C. canephora*, which was nearly globose – having almost equal maximum longitudinal length and major radial axis length. The two varieties of *C. liberica* were both ellipsoid, but *C. liberica* var. Liberica was longer than wide compared to C. *liberica* var. Dewevrei. The results obtained for fruit measurements are consistent with the observations of Tao and Taylor (2011), which noted the same shapes for these varieties.

Geomorphometric Data

C. arabica was the most distinguishable *Coffea* variety based on berry shape, having the least misclassifications using CVA–MANOVA. The large intraspecific variability of *C. arabica* is mainly due to its autogamous and tetraploid nature (Charrier & Berthaud 1985). By being self-fertilizing, it promotes homozygosity in its gene loci, allowing for more dissimilarity between other species (Acquaah 2012).

C. canephora can be distinguished from other *Coffea* varieties based on its longitudinal berry shape. Its radial berry shape, however, is insufficient to separate it from other varieties. Misclassifications in radial berry shape were mostly between *C. canephora* and *C. liberica* varieties. *C. liberica* var. Liberica can be differentiated using its longitudinal berry shape. On the other hand, its radial berry shape is most similar to *C. arabica*, as seen from its misclassifications in the CVA plot.

Lastly, *C. liberica* var. Dewevrei – in all shape analysis – had misclassifications in *C. canephora* and *C. liberica* var. Liberica and as such, cannot be distinguished based on its shape. To differentiate this variety, metric data must be used instead.

In terms of shape, several misclassifications were observed between *C. canephora* and the two *C. liberica* varieties. *C. arabica* was the only species that could be distinguishable in all its shape analyses. These results could be attributed to their ploidy. *C arabica* is the only tetraploid *Coffea* species while *C. canephora* and *C. liberica* are both diploid (Herrera *et al*. 2002). *C. canephora* and *C. liberica* may have clustered together due to their similar ploidy while *C. arabica*, being the only tetraploid, formed its own cluster.

CONCLUSIONS

In differentiating the four varieties of *Coffea* (*C. arabica, C. canephora, C. liberica* var. Liberica, *C. liberica* var. Dewevrei), a combination of metric and geomorphometric characteristics must be used. Among all metric characters, the length of the radial major axis was the most important in differentiating their berries. In their geomorphometric analysis, the berry shape in longitudinal view was able to distinguish varieties with the least misclassifications. *C. arabica* can easily be distinguished from other varieties on the basis of its berry shape and length of its radial minor axis. *C. canephora* can be differentiated by its longitudinal berry shape and length of its radial major axis. *C. liberica* var. Liberica is recognizable by its longitudinal berry shape and also the length of its radial major axis. Lastly, *C. liberica* var. Dewevrei can be identified solely by its metric character, the radial major axis.

ACKNOWLEDGMENT

The authors would like to thank the Bureau of Agricultural Research – Department of Agriculture for the research funding. This work is part of the Project entitled "Characterization of Philippine Coffee Varieties" under Study 4. Genetic Variation in Philippine Coffee using Multivariate Morphometry and Geomorphometric Analysis. Fruit samples were obtained from the germplasm collection of the National Coffee Research Development and Extension Center of the Cavite State University in Indang, Cavite. Laboratory processing and analysis was done at the Plant Genetics and Cyanobacterial Biotechnology Laboratory, Institute of Biology, University of the Philippines in Diliman, Quezon City.

REFERENCES

ACQUAAH G. 2012. Principles of plant genetics and breeding. 2nd ed. New York: John Wiley and Sons, Ltd. p. 97–112.

ADAMS DC, ROHLF FJ, SLICE DE. 2004. Geometric morphometrics: Ten years of progress following the 'revolution'. Italian Journal of Zoology 71: 5–16.

ANTHONY F, BERTRAND B, QUIROS, O, WILCHES A, LASHERMES P, BERTHAUD J, CHARRIER A. 2001. Genetic diversity of wild coffee (*Coffea arabica* L.) using molecular markers. Euphytica 118: 53–65.

CAO EP, SANTOS DC, RAMOS LP, SANTOS BS, QUILANG JP, MOJICA RM. 2014. Molecular and morphological differentiation among *Coffea* (Rubiaceae) varieties grown in the farms of Cavite province, Philippines. Philippine Science Letters 7(2): 387–397.

CHARRIER A, BERTHAUD J. 1985. Botanical classification of coffee. Westport, Connecticut: The AVI Publishing Company.

CHONG SML, SANTOS DC, CAO EP. 2014. Pollen morphometrics of four coffee (*Coffea* sp.) varieties grown in the Philippines. Philippine Journal of Crop Science 39(3): 1–7.

CLARKE KR. 1993. Non-parametric multivariate analyses of changes in community structure. Australian Journal of Ecology 18: 117–143.

DAVIS AP, GOVAERTS R, BRIDSON D, STOFFELEN P. 2006. An annotated taxonomic conspectus of the genus *Coffea* (Rubiaceae). Botanical Journal of the Linnean Society 152: 465–512.

HAMMER O, HARPER DAT, RYAN PD. 2001. PAST: Paleontological statistics software package for education and data analysis. Palaeontologia Electronica 4(1): 9p. http://palaeoelectronica.org/2001_1/past/issue1_01.htm

HERRERA J, COMBES C, CORTINA H, ALVARADO G, LASHERMES P. 2002. Gene introgression into *Coffea arabica* by way of triploid hybrids (*C. arabica* x *C. canephora)*. Heredity 89(6): 488–494. doi:10.1038/sj.hdy.6800171.

MAURIN O, DAVIS AP, CHESTER M, MVUNGI EF, FAKIM Y, FAY MF. 2007. Towards a phylogeny for *Coffea* (Rubiaceae): Identifying well-supported lineages based on nuclear and plastid DNA sequences. Annals of Botany 100(7): 1565–83.

MISHRA MK, SLATER A. 2012. Recent advances in the genetic transformation of coffee. Biotechnology Research International. p. 1–17.

ROHLF FJ, SLICE D. 1990. Extensions of the procrustes method for the optimal superimposition of landmarks. Systematic Zoology 39(1): 40–59.

TAO C, TAYLOR CM. 2011.*Coffea* Linnaeus, Sp. Pl: 1:172. 1753. Flora of China 19: 90–92.

ZELDITCH M, SWIDERSKI D, SHEETS H. 2012. Geometric morphometrics for biologists: A primer, 2nd ed. USA: Academic Press. 488p.