Potential Use of Morphological Characteristics in Evaluating Natural Variation of "Barako" Seedlings

Julienne Davis V. Aguilar, Denise Lindsey T. Kiok, Bernice Nicole I. Macalinao, Ritchelle C. Ting and Khiyomi Hannah B. Uy
De La Salle University Integrated School, Manila

Abstract: The Philippines is one of only three countries able to produce Coffea liberica and grow two other species of coffee, Coffea canephora, Coffea arabica. Coffea liberica, or "Barako," is the least cultivated of coffee species. Its unique taste has high potential to grow in the market if production was increased. However, the 2020 Taal Volcano eruption resulted in thousands of damaged Barako trees in 2020 and 2021. As new seedlings are produced, the question of variety among planting material comes up. Farmers need a method to evaluate natural differences of current Barako trees, to answer whether seedlings from farms cultivating Barako are naturally different. This research aims to determine if the morphological characteristics of cultivated seedlings from different farms can be classified through image and statistical analysis. Basic knowledge on how varieties—which produce different flavors, aroma, and market value of coffee—differ among farms is needed for a strategy to increase the number of seedlings. This study's initial data set indicated statistically significant differences in the average seedling height per node and the leaf area per length of 31, 10-month old coffee seedlings from two different farms, grown in a common environment. The process may be developed further for use in evaluating natural variation among C. liberica as seedlings.

Key Words: coffee; barako seedlings; Coffea liberica; data analysis; morphological characteristics.

1. INTRODUCTION

Coffea liberica, commonly known as "Barako" coffee, is a species of coffee plant under the Rubiaceae family made known by William Hiern in 1876. *C. liberica* only accounts for about 1% of the Philippine's coffee production, with *Coffea canephora* as the largest (Philippine Statistics Authority, 2019). Farmers replace their Barako with Robusta and short, hybrid coffee trees because of the surge in demand for instant coffee. With this problem, Barako has been facing difficulties finding its way onto the menus of mainstream coffee (Kapeng Barako, 2018).

The Philippines has a competitive advantage in the global production of *Coffea liberica*, being one of three countries—besides Ethiopia and Malaysia—that can produce *Coffea liberica* (Barako). *C. liberica* accounts for about 3-4% of overall coffee production in the world (Wallengren, 2018). With the current growing market for specialty coffee, *C. liberica* is finding new markets. The species, however, is on the verge of extinction, given very few farms exist. The Taal eruption during early 2020 in Batangas, a province famous for its production of Barako coffee, has further diminished the supply of Barako.

The value of coffee is dependent on its qualities. Variation in harvest quality results in price fluctuations where certain species are sold at higher or lower market prices. The same is true in many different plants, one of which is mangoes (Department of Agriculture, 2019). Certain varieties of mango are sold at higher prices. Farmers, sellers, and buyers base prices on morphological characters that distinguish types and product qualities. For coffee, the market of coffee beans and coffee seedlings is determined in the same way. Phenotypic characters of morphology and quality distinguish varieties grown in different areas. There are more long-term benefits for farmers and plant breeders to use physical characteristics because these are more practical (Kordrostami & Rahimi, 2015). Differences like seedlings (i.e., species, variety, or stock) may equate to differences in other qualities such as tastes, aroma, etc.

Since the Barako coffee has been in demand in the coffee market, the effort to increase seedlings of Barako production will benefit from information on knowing natural differences in Barako stocks. If farmers cultivate different stocks, it would make sense to tag each differently and see if there are corresponding differences in taste, disease resistance, yield, etc. Hence, being able to differentiate the species of Barako based on the morphology of seedling and leaf characters would be necessary. The gold standard for this would be an analysis of genetic markers, but this would not be feasible for farmers. The process requires much training, equipment, and time, which are not necessarily available to farmers (Rahman et al., 2009).%

Image analysis programs and statistical tools were used to figure out morphological clusters formed from the data gathered. ImageJ is an image processing program capable of calculating area and pixel values of user-defined selections—measuring distances and angles (Bankhead, 2014). It can quantify or measure the plant's visible traits, such as leaf morphology, to detect the link between its genome and its physiological characteristics (Kokorian et al., 2010). The study used 41 *Coffea liberica* (Barako) seedlings to undergo image analysis to differentiate its morphological characteristics, obtained from two farms grown in a controlled environment. This study was done with the approval of the research faculty responsible for the Senior High School students of DLSU. DNA markers were not used to verify whether the Barako seedlings are of different varieties.

The study can be of great help in assisting farmers, sellers, buyers of coffee in quickly determining *Coffea liberica* varieties. This can help in strategies to increase production of Barako, and hopefully the income as well of those involved in the industry; to encourage them to produce and preserve Barako. Doing so can prevent its decrease, meet the demand for new flavors in the coffee market, and improve the qualities of *C. liberica* varieties in the future.

2. METHODOLOGY

2.1. Sample Collection

A total of 41 ten-month-old *C. liberica* seedlings from 2 farms in Sarawak, Malaysia, were used as data for the morphological characteristics—31 for leaf area per leaf length and 41 for seedling height per seedling node. They were grown together in a controlled environment.

2.2 Data Collection

A camera was used to take images of these seedlings with a meter stick beside them to gather the height and number of nodes (Figure 2.2.1). Four to five leaves were collected by taking images in a parallel manner (Figure 2.2.2). The previously gathered 41 seedlings were named and segregated by category. Data collected from the coffee seedlings and their leaves are added to Microsoft Excel for data analysis and graphs.

Figure 2.2.1 *Seedling with the measuring stick*

Figure 2.2.2 *Sample Leaf with Ruler*

2.3. Data Analysis

2.3.1. Mean and Standard Deviation

The mean was taken from the raw data of the four main characteristics in the seedlings to see the overall view of the seedlings in groups "Farm 1" and "Farm 2" and as a whole, "All Samples." Standard deviation was also used to determine the closeness of the variables. A higher value for the standard deviation tells that the data is more spread out, while a smaller standard deviation tells the proximity of the data.

2.3.2. Histogram

The histogram was used to determine if two groups can be observed based on the standardized value of the morphological characteristics of the seedlings. If the type of histogram will be skewed to the left or skewed to the right, the data has a high correlation, and that the two groups are possibly overlapping each other. If it is bimodal, it shows that there are indeed two groups due to a low correlation.

2.3.3. Scatterplot

The morphological characteristics of the seedlings and their respective leaves were compared in scatter plots to determine their correlation between each other. The correlation identifies if two groups exist and if these variables are applicable in the differentiation of morphology in *Coffea liberica* seedlings. High correlation is present if the data points are clustered together and have a large R^2 value. Low correlation within scatter plots is present if data points are scattered in a broader range, where it is possible to view two groups within the plot area. Indicated trendline acts as a divider to check for the visible possibility of two groups between the data points, along with the cluster of the data. Standardized values are not found in the scatter plots.

2.3.4. Paired T-Test

A Paired t-test was applied to determine whether there is a significant difference between the variability of the two farms based on the standardized values of Seedling Height per Seedling Node and Leaf Area per Leaf Length. When the T-test value is greater than the value from the distribution table, then the null hypothesis that the paired population is equal is rejected; if otherwise, accepted.

3. RESULTS AND DISCUSSIONS

3.1. Table of Means

The standardized values for both parameters, seedling height per nodes and area per length, shows that the means from Farm 1 is less than the mean for all the samples, while the means from Farm 2 is greater than the means for all samples. The table below (Table 3.1.1) showed that the mean for the seedlings in Farm 2 is taller than Farm 1. The values of the standard deviation of Farm 2 are more significant in most characteristics, except for the number of nodes and height per number of nodes. Higher standard deviation results in a more considerable variation of values, making it harder to differentiate due to a broader range of values.

Table 3.1.1 *Table of Means*

	Farm 1		**Farm 2**		**All**	
	Mean	**SD**	**Mean**	**SD**	**Mean**	**SD**
Height (cm)	19.2	5.4	26.9	8.03	21.5	8.7
No. of Nodes	5.8	1.4	6.3	1.2	5.9	1.4
Length (cm)	12.3	2.9	14.7	3.5	13.6	3.5
Area (cm^2)	43.8	20.7	63.1	31.7	54.7	29.1
Height per No. of Nodes	3.2	1.05	4.5	1.02	3.8	1.8
Area per Length (cm)	3.5	0.9	4.2	1.1	3.8	1.1

3.2 Scatter Plot

3.2.1. Leaf Length vs. Leaf Area

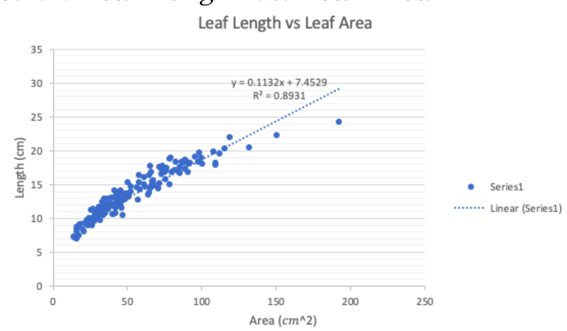

Figure 3.2.1.1 *Scatter Plot of Leaf Length vs. Leaf Area*

The graph (Figure 3.2.1.1.) shows a cluster with a high R^2 value of 0.8931, which indicates a high correlation. Due to this indication, these data sets cannot be used separately and must be standardized into "Leaf Area per Leaf Length." While it is possible to use one, standardization is recommended due to the high correlation. Leaf length was used to standardize leaf area.

3.2.2. Seedling Height vs. Seedling Nodes

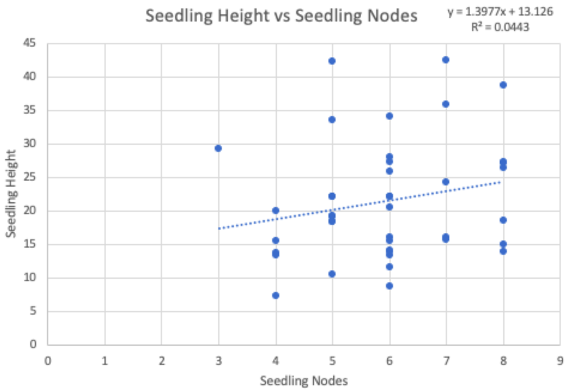

Figure 3.2.2.1 *Scatter Plot of Seedling Height vs. Seedling Nodes*

Data points (Figure 3.2.2.1.) between seedling height (y) and the number of nodes (x) present a wider spread of plot area covered, with less cluster between points. This low correlation is also shown with a smaller R^2 value of 0.0396, indicating that these variables have very high variability in the data. The separation gives them a significant variation, and it is difficult to compare the sets of data unless they are standardized into "Seedling Height per Seedling Nodes." This resolves the wide variation that would otherwise make it challenging to compare height with nodes. It is also possible to use one variable, but the variation makes it difficult to choose. Therefore, a standardized value makes it easier to have both variables for the two groups of *C. liberica*.

3.2.3. Comparison of data points from results of Farm 1 and Farm 2

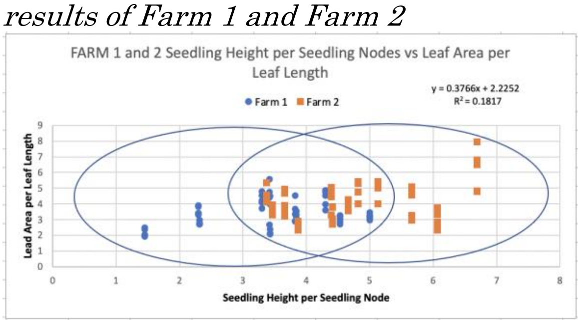

Figure 3.2.3.1 *Comparison of data points from results of Farm 1 and Farm 2*

Once the overall variables were standardized and combined, they appear to form two overlapping groups (Figure 3.2.3.1). Some variables from Farm 1 and Farm 2 have merged in the middle of the group (between the range of 3 to 5). Combining all characters measured produces a graph that shows groups. The two groups formed are Farm 1 (blue points) located on the leftmost and Farm 2 (orange points) on the rightmost. They are not enough to completely separate them into varieties. It is possible to investigate further the properties that would make the groups distinct by introducing more seedlings, characters, and farms.

3.3 T-test

The ability to determine differences in two groups using all the morphological characteristics measured is further emphasized through the T-test.

3.3.1. Seedling Data for Height per Nodes

Table 3.3.1.1 *T-test data for Seedling Height per Nodes*

	Farm 1	Farm 2
Mean (X)	3.1	4.6
Sample size (N)	14	13
Standard deviation (S)	1.4	1.05
Paired T-Test Value for Seedling Height per Nodes	3.0802	

When the seedling height by nodes of 14 seedlings from Farm 1 ($X_A = 3.1$, $S_A = 1.4$) and the data of the other 13 seedlings from Farm 2 ($X_B = 4.6$, $S_B = 1.05$) were compared, the resulting t-test value score was $t(12) = 2.306$, $p < .05$. This shows differences between the means, as the p-value or threshold acquired from the t-test table (2.306) is lower than that of the t-test score result (3.0802). Therefore, the null hypothesis stating that there is no significant difference between the two farms is rejected.

3.3.2. Leaf Data for Area per Length

Table 3.3.2.1 *T-test data for Leaf Area per Length*

	Farm 1	Farm 2
Mean (X)	3.4	4.1
Sample size (N)	55	63
Standard deviation (S)	0.9	1.09
Paired T-Test Value for Leaf Area per Leaf Length	3.7892	

Comparing the leaf area by length of 55 leaves from 11 seedlings from Farm 1 ($X_A = 3.4$, $S_A = 0.9$) with the data of the other 63 leaves coming from 13 seedlings of Farm 2 ($X_B = 4.1$, $S_B = 1.09$); the calculated t-test value score was $t(54) = 3.7892$, $p < .05$, indicating significant differences between the means. As the p-value or threshold acquired from the t-test table (2.0154) is lower than that of the t-test score result (3.7892), the null hypothesis that there is no significant difference between the two farms is rejected.

4. CONCLUSION AND RECOMMENDATION

The four parameters: seedling height, seedling nodes, leaf length, and leaf area, are best used as standardized values since it partially separates the groups. These then qualify as characteristics needed to carry out a paired t-test calculation—a statistical analysis done to determine differences between two variables of the same subject. After the seedling samples were classified into two groups, there was a significant difference between them; as two overlapping groups formed from the scatter plot (Figure 3.2.3.1). Thus, these indicate a potential of differentiation by using standardized morphological characteristics of seedlings grown in a controlled environment.

For future studies, it is recommended to get more seedling samples to test and gather more data to distinguish seedlings from various farms. Adding more morphological characteristics to separate further the groups, such as leaf apex, leaf margin, leaf venation, phyllotaxy, etc, is highly recommended. As suggested by the DLSU SHS Research Congress committee (personal communication, 2021; Nakano, as cited in KPU Pressbooks, 2020), morphological identification can also use flowers since it is said to be the part less affected by growth conditions. A larger population and other coffee seedlings will perhaps allow larger variation between *C. liberica* seedlings. Other farms can also be included, and their seedlings can be tested for variation or differentiation of their morphological characteristics, both with the existing parameters and additional ones.

5. ACKNOWLEDGMENTS

This research would not have been made possible without the devotion and assistance of their research adviser, Dr. Ma. Carmen Ablan-Lagman, Ph.D., who was also the provider of the coffee seedlings used in this research. For this, the group would like to express their deepest gratitude. She has shared her knowledge and experiences with the whole team, inspiring them to do better in all aspects of research. They would also like to thank Dr. Chona Camille Vince Cruz-Abeledo, and Engr. Michael Manguerra, their Practical Research 1 and 2 professors, respectively, for teaching them the basics of research.

6. REFERENCES

Bamber, P., Daly, J., & Gereffi, G. (2017, April). The Philippines in the Coffee Global Value Chain. http://industry.gov.ph/wp-content/uploads/2017/08/The-Philippines-in-the-Coffee-Global-Value-Chain.pdf

Bankhead, Peter. (2014, June). Analyzing fluorescence microscopy images with ImageJ. https://www.researchgate.net/publication/260261544_Analyzing_fluorescence_microscopy_images_with_ImageJ

Cao, E., Constantino-Santos, D. M., Ramos, L. A., Santos, B., Quilang, J., & Mojica, R. (2014, November 25). Molecular and morphological differentiation among Coffea (Rubiaceae) varieties grown in the farms of Cavite province, Philippines. Philippine Science Letters, 7(2), (pp. 387-397).

Department of Agriculture. (2019, June 14). Philippine mango industry roadmap. https://www.da.gov.ph/wp-content/uploads/2019/06/Philippine-Mango-Industry-Roadmap-2017-2022.pdf

Kapeng Barako: The Endangered Philippine Liberica. (2018, November 27). Prima Donna Life. https://theprimadonnalife.com/lifestyle/kapeng-barako-the-endangered-philippine-liberica/

Kokorian, J. (2010, October 29). An ImageJ based measurement setup for automated phenotyping of plants. In G. Polder, J. J. B. Keurentjes, D. Vreugdenhil, & M. O. Guzman (Eds.), Proceedings of the ImageJ User and Developer Conference, Luxembourg, Luxembourg. (pp. 178–182). https://library.wur.nl/WebQuery/wurpubs/398563

Kordrostami, M., & Rahimi, M. (2015, September 15). Molecular markers in plants: concepts and applications. Genetics in The 3rd Millennium, 13, (pp. 4024-4031).

Nakano, M. (2020, February 13). Introduction to Plant Morphology [Red Seal Landscape Horticulturist Identify Plants and Plant Requirements (F2 - 1&2)]. KPU Pressbooks. https://kpu.pressbooks.pub/plant-identification/chapter/introduction-to-plant-morphology/

Philippine Statistics Authority. (2020). Major non-food and industrial crops quarterly bulletin, October-December 2020. https://psa.gov.ph/non-food/coffee.

Rahman, M., Zafar, Y., & Paterson, A. (2009, February 27). Gossypium DNA markers: Types, numbers, and uses. https://doi.org/10.1007/978-0-387-70810-2_5.

Wallengren, P. B. M. (2020, May 10). Coffee of the day: The Elusive Philippines Liberica Barako Coffee Bean. Spilling the Beans. https://philcoffeeboard.com/coffee-of-the-day-the-elusive-philippines-liberica-barako-coffee-bean/

Wintgens, Jean Nicolas (2004, July 27). Coffee: Growing, processing, sustainable production (A Guidebook for Growers, Processors, Traders, and Researchers). https://doi.org/10.1002/9783527619627.