© NIAB 2019
ISSN 1479-2621
Plant Genetic Resources: Characterization and Utilization (2019) **17**(6); 480–487
doi:10.1017/S1479262119000285

An objective method of shape descriptor state establishment using elliptic Fourier analysis (EFA)

Renerio P. Gentallan Jr.[1]*, Nestor C. Altoveros[1], Teresita H. Borromeo[1], Leah E. Endonela[2], Fiona R. Hay[3], Antonio G. Lalusin[1], Consorcia E. Reaño[1] and Yosuke Yoshioka[4]

[1]*Institute of Crop Science, College of Agriculture and Food Science, University of the Philippines Los Baños, College, Laguna, Philippines,* [2]*Institute of Plant Breeding, College of Agriculture and Food Science, University of the Philippines Los Baños, College, Laguna, Philippines,* [3]*Department of Agroecology, Aarhus Universit, Forsøgsvej 1, 4200 Slagelse, Denmark and* [4]*Faculty of Life and Environmental Sciences, University of Tsukuba, Ibaraki, Japan*

Received 16 April 2019; Accepted 23 August 2019 – First published online 2 October 2019

Abstract

This research attempts to systematically establish shape descriptor states through elliptic Fourier analysis (EFA) using pili (*Canarium ovatum* Engl.) kernel as a model. Kernel images of 53 pili accessions from the National Plant Genetic Resources Laboratory (NPGRL), University of the Philippines Los Baños were acquired using VideometerLab 3. Shape features, such as roundness, compactness and elongation, were extracted from the images. Shapes outlines were characterized using elliptic Fourier coefficients calculated from SHAPE version 1.3 software. Principal component analysis and cluster analysis were used to elucidate clusters representing the shape descriptor states. The first principal component accounts for the variation in length to width ratio; whereas, the second and third principal components explain the variation in the location of the widest portion and the truncation of the apex and base of the kernel, respectively. Cluster analysis separated the different accessions into six distinct clusters at 0.04 Euclidian distance. Six descriptor states, narrowly elliptic, elliptic, widely elliptic, ovate, obovate and lance-ovate, were characterized from the shape outlines and visualized through R's shape on r package. The discrimination between clusters was validated through MANOVA and LDA with 95% correct classification. The Fourier coefficients were also able to represent the variation observed from the physical properties of shape. The method may be used in establishing shape descriptors of all plant parts of all crop species.

Keywords: descriptor states, elliptical Fourier analysis, image analysis, plant phenotyping, shape

Introduction

The creation and use of a crop descriptor list, an assembly of varying plant characters, has been promoted by Bioversity International to characterize and evaluate plant genetic resources. Each descriptor, a trait, has corresponding 'descriptor states' which are distinguishable states of expression of the specified trait that potentially represent all the observed trait variations of the entire species (Bioversity International, 2007). The clear and definite representation provided by the descriptor list facilitates ease in universally describing plant germplasm collections for their subsequent utilization. As of 2018, there are 102 descriptor lists of various crops published by Bioversity International; however, many crop species have no established crop descriptors, limiting their formal characterization. Limitations include the use of subjective assessment in

*Corresponding author. E-mail: rpgentallan@up.edu.ph

https://doi.org/10.1017/S1479262119000285 Published online by Cambridge University Press

establishing these lists, as highlighted by Gentallan *et al.* (2019) in proposing a systematic protocol of colour descriptor state formulation. Particularly for traits with irregular shapes, there is difficulty in accurately identifying clear and definable states due to ambiguity in the process of representation. To address this, several image-based phenotyping techniques have been used to characterize the shapes or plant organs such as fruits (Wycislo *et al.*, 2008), flowers (Neto *et al.*, 2009) and seeds (Cervantes *et al.*, 2016). These include the use of elliptic Fourier analysis (EFA) that analyses closed contours immaterial of the size and configuration (Kuhl and Giardina, 1982). This method describes mathematically the overall shape of an object by transforming outline information into Fourier coefficients. Plant characterization using this method was first attempted by White *et al.* (1988) who analysed the leaf shape of Betula. Since then, several studies of plant shape characterization using EFA have been published, for example, such as in the case of petal shape in Japanese primrose (Yoshioka *et al.*, 2004), grain shape in rice (Iwata *et al.*, 2015) and fruit shape in persimmon (Maeda *et al.*, 2018). Moreover, in defining qualitative descriptors, such as shape and colour where subjectivity often occurs, the 'use of image analysis eventually to convert qualitative data to quantitative data' has been encouraged (Bioversity International, 2007). In spite of this, there are no reports on the application of image-based techniques, particularly EFA, to systematically establish shape descriptor states.

Pili (*Canarium ovatum* Engl.), a native tree nut of the Philippines with great economic potential (Coronel, 2011), shows kernel shape variation but has no established descriptor list. Thus, this study aimed to objectively identify the shape descriptor states by elliptic Fourier descriptors using pili (*C. ovatum* Engl.) kernels as a model.

Materials and methods

Plant materials

The plant materials consisted of 53 pili accessions from the Institute of Crop Science, and the National Plant Genetic Resources Laboratory of the Institute of Plant Breeding, College of Agriculture and Food Science, University of the Philippines Los Baños. Each accession was represented by 10 samples. To avoid shape and size variation due to varying seed moisture contents (mc), samples were air dried to approximately 5% mc prior to image acquisition.

Image acquisition

Kernel images were captured using the Videometerlab3 set up (Videometer A/S Hørsholm, Denmark). To ensure standardized and comparable images, the set up was calibrated for colour, geometry and self-illumination using standard uniform bright, uniform dark and geometric discs. Samples were then placed on a blue background disc, and images with a size of 2056 × 2056 pixels and a resolution of ∼45 µm/pixel were captured using the set up.

Computation of elliptic Fourier coefficients

A modified method developed by Yoshioka *et al.* (2004) to analyse shape variation through EFA was used. Prior to analysis, the images were formatted into 24-bit bitmap (*.bmp) using SigmaScan®Pro 5.0 software. The converted images were chain-coded and normalized using Shape version 1.3 software (Iwata and Ukai, 2002). Initially, using the ChainCoder of Shape ver. 1.3, closed contours of the kernel and seed were obtained from the binary images. The contours were then chain-coded using the method proposed by Freeman (1974). The pixel contour of the digital images served as the discrete approximation of a continuous curve and was recorded as chain codes to describe the geometrical information on outlines ranging from 0 to 7. From the chain codes, the elliptic Fourier coefficients of the n^{th} harmonic (a_n, b_n, c_n, d_n) were computed to describe the shape variations in the pili accessions. Twenty (20) harmonics were used as they were considered to be sufficient in describing shape variations for morphological analysis (Yoshioka *et al.*, 2004; Sayinci *et al.*, 2014). Then, the computed Fourier coefficients were standardized for size, orientation and starting point variations using the method of Kuhl and Giardina (1982).

Image-based characterization of pili kernel physical properties

To validate the descriptor formulated states, the projected area (A, mm^2), length (L, mm), width (W, mm) and perimeter (P, mm) of each kernel and seed were determined using VideometerLab 3 software. Subsequently, from the image-based measurements, roundness (R), pointness (Po), elongation (El), eccentricity (Ec) and compactness (C) were calculated using the equations in online Supplementary Table S1.

Principal component analysis, cluster analysis and visualization

From the elliptic Fourier coefficients of the sample outlines, principal component analysis was performed using the variance–covariance matrix to obtain fewer, uncorrelated variables that would effectively summarize the variation observed in the coefficients (Rohlf and Archie, 1984). The data were analysed in R Version 3.4.3. Using Perl Strawberry version 5.26.1.1, normalized elliptic Fourier

https://doi.org/10.1017/S1479262119000285 Published online by Cambridge University Press

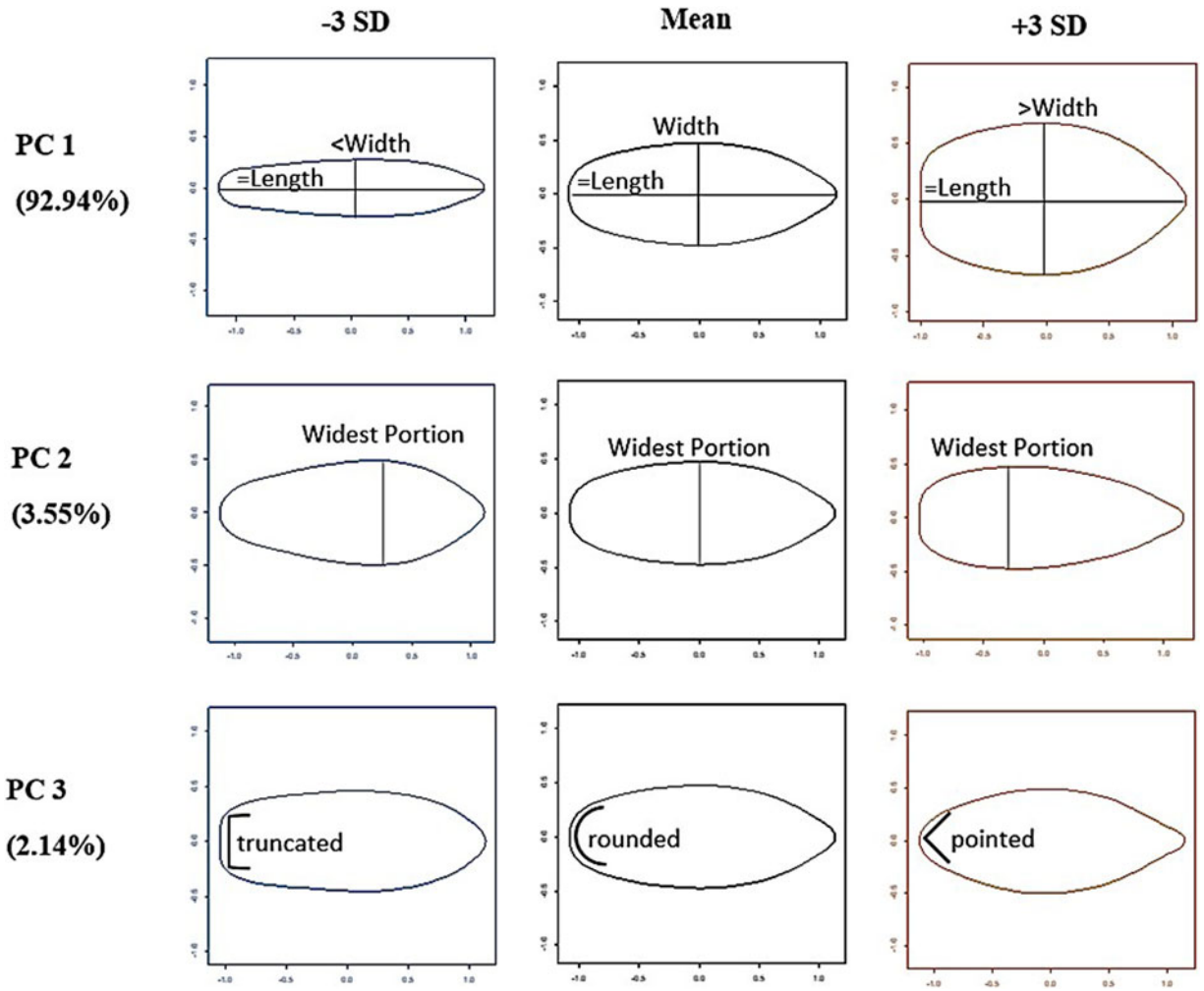


Fig. 1. Shape variations explained by the first three principal components; SD indicates standard deviation.

descriptors were converted to dat format, for R software readability. The scores of the components were computed using the package SHAPE on R (Iwata and Ukai, 2002) and used as shape characteristics for succeeding analyses. The dendrogram was created using unweighted pair group method with arithmetic means (UPGMA) and sequential, agglomerative, hierarchical and nested clustering parameters (SAHN) program of Numerical Taxonomy and Multivariate Analysis System (NTSYS)-pc version 2.1 (Rohlf, 2002). The dendrogram was cut into clusters based on the variance within and between clusters. As a general rule, the dendrogram was statically cut at the point where the minimum number of clusters satisfies the following conditions: variance within clusters is ⩽11% and any additional cluster will entail <2% decrease in the variance within clusters. This will minimize variance within clusters and maximize any additional benefit of added variance between clusters by adding an additional cluster. The variances were computed using XLSTAT 2016 (Addinsoft, Inc., Brooklyn, NY, USA). From the generated clusters, shape Fourier coefficients were averaged within clusters. The averaged values per cluster were used to represent a descriptor state. The averaged cluster values for the principal component values were visualized using the draw functions of the package SHAPE on R (Iwata and Ukai, 2002).

Analysis of variance (ANOVA), multivariate ANOVA (MANOVA) and pair-wise mean comparisons

To examine the differences in physical properties among accessions, ANOVA at 5% level of significance was performed for each physical property. If at least one treatment was considered to be different, an LSD test was conducted to distinguish the differences between treatments. For the kernel outline differences between clusters, MANOVA at 5% level of significance was performed. Upon observing

https://doi.org/10.1017/S1479262119000285 Published online by Cambridge University Press

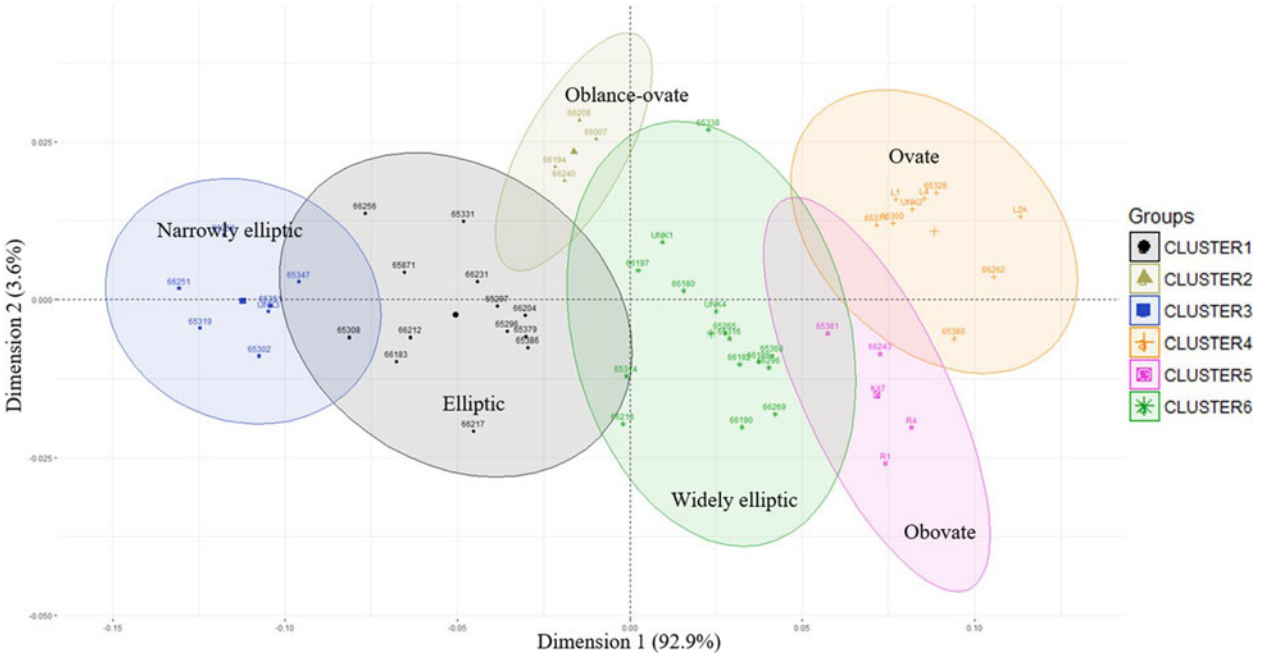


Fig. 2. Plot of component loadings obtained from PC1 and PC2 for the 53 pili accessions; clusters are enclosed by concentration ellipses.

a significant difference between clusters, pairwise comparisons from Fisher distance and *P* values were calculated at the same level of significance.

Linear discriminant analysis (LDA)

To validate further the descriptor states, a separate set of 20 kernel outlines per cluster was evaluated using LDA. Per cent correct classification rates between clusters were used to gauge the accuracy of discriminating individual outline shapes based on the clusters. LDA was implemented using XLSTAT 2016 (Addinsoft, Inc.).

Results

The first principal component accounted for 92.94% of the total variation; 98.62% of the total variance could be explained by the first three components (Figs. 1 and 2). It was determined that the first principal component accounts for the variation in length to width ratio; whereas, the second and third principal components explain the variation in the location of the widest portion and the truncation of the apex and base of the kernel, respectively (Fig. 1).

From the variance within and between clusters with an increasing number of classes, it was determined that five clusters will impose 9.40% variance within clusters; furthermore, an additional class will result only to an additional decrease of 1.37%. Hence, the dendrogram was truncated to produce six clusters corresponding to six descriptor states for kernel shape – narrowly elliptic, elliptic, widely elliptic, ovate, obovate and lance-ovate (Figs. 3 and 4).

With a low Wilk's Lambda of 0.033, significant differences among the six clusters were observed based on the kernel outlines (Table 1). The pair-wise comparisons between the six clusters displayed significant levels of difference indicating an effective classification based on kernel outlines. From the pre-classified kernel outlines, the discriminant analysis showed 95.00% correct classification (Table 1). The highest classification was observed for cluster 3 with 100% and the lowest was observed at cluster 6 with 90.00%. Furthermore, highly diverse accessions were observed for all the physical shape parameters with an *H* of >0.67 (online Supplementary Table S2).

Discussion

Clusters 3, 1 and 6 represent the elliptical series of shapes – narrowly elliptic, elliptic and widely elliptic, respectively (Figs. 3 and 4). Showing non-overlapping points across PC1, these three clusters are highly influenced by the first principal component and are easily discriminated through the principal coordinate analysis of the first two components (Fig. 2). The states match with the described shape classification of the elliptic series of shapes by Simpson (2010). The three states are considered elliptic as the widest part is located at the midpoint of the kernel and the margins are curved; however, variation in length to width ratio is observed. From narrowly to widely elliptic, the length to width

https://doi.org/10.1017/S1479262119000285 Published online by Cambridge University Press

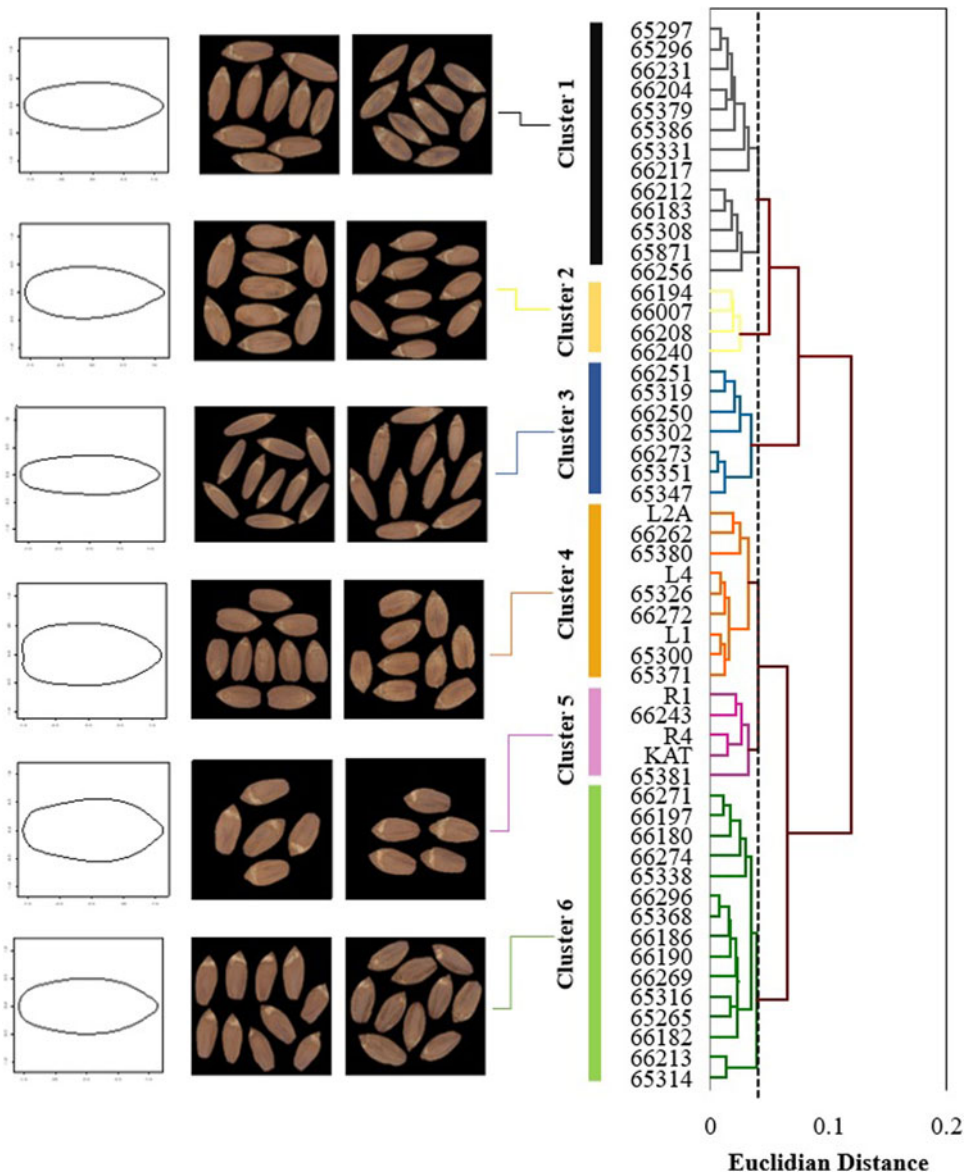


Fig. 3. Dendrogram of 53 pili accessions from average elliptical Fourier coefficients of shape outlines.

ratio decreases. The three states were well represented in the germplasm collection with 15, 13 and seven accessions having widely elliptic, elliptic and narrowly elliptic kernel shapes, respectively (Fig. 4).

Cluster 2 is characterized to have an oblance-ovate shape, also known as inverted lachryform or tear-shaped. This shape has curved margins that are widest near the base and the length to width ratio is between 3:1 and 2:1 (Simpson, 2010), indicating a high length to width ratio. This cluster is highly influenced by the second principal component with the movement of the widest point from the centre to near the apex (Fig. 1). Thus, accessions from this state were completely separated at PC2 in the principal coordinate analysis (Fig. 2). Inverted lachryform shapes, however, are only represented by four accessions in the germplasm collection (Fig. 3).

Clusters 4 and 5 represent ovate and obovate shapes, respectively. Having longer widths, they are differentiated in the first principal component by their relatively lower length to width ratio ranging from 2:1 to 3:2 (Simpson, 2010). These shapes are differentiated from each other in the second principal component. Obovate-shaped accessions have the widest point near the apex, whereas the ovate-shaped accessions have widest points near the base as shown in the second principal component (Fig. 2).

Using the validation set, a relatively high correct classification rate was observed compared to the 34.9% correctness in classifying hazelnut outlines based on cultivars by

https://doi.org/10.1017/S1479262119000285 Published online by Cambridge University Press

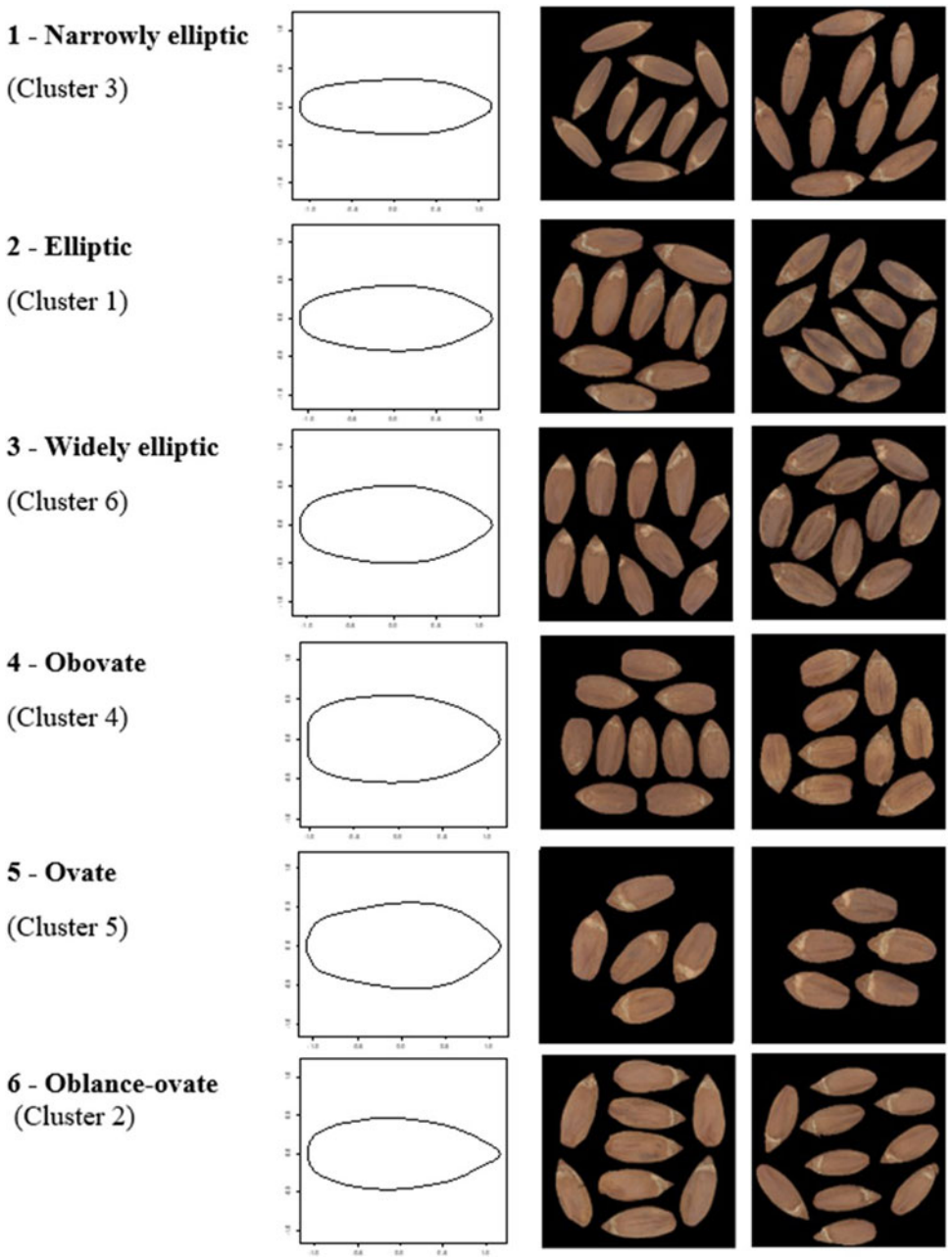


Fig. 4. Pili kernel shape descriptor states.

Sayinci *et al*. (2014). Low classification rates on a per accession basis may be due to similarities in outline characteristics between accessions. Hence, the method provides an effective means of summarizing shape similarities and variations for subsequent representation.

Given the high diversity, roundness values ranged from 2.1 to 3.4. A roundness of 1.0 denotes a perfect circle while larger values indicate oblong and non-circular objects. This indicates that the germplasm accessions observed were oblong to non-circular as verified by the descriptor formulated states. Narrowly elliptic descriptor states and those with higher length to width ratio have higher roundness values compared with descriptor states with lower length to width ratio. This is exemplified by accession 66183 having narrowly elliptic kernels with the lowest score for roundness. On the other hand, L4, with the highest roundness value was classified to be obovate. For eccentricity, similar trends can be observed, as a circle will have a value of zero, whereas a straight line will have a value of 1 (Birchfield, 2016). Pointedness, on the other hand, defines how subulate is the basal part of the kernel. Highly pointed accessions will have higher pointness values. This parameter is

https://doi.org/10.1017/S1479262119000285 Published online by Cambridge University Press

Table 1. Differences between clusters based on kernel outlines and their per cent correct classification from discrimination analysis of individual kernel outlines

(A) MANOVA, Wilk's Lambda (computed in XLSTAT 2016)

Effect	Value	df_1	df_2	*F*-crit	*F*-obs	*p*-value
Clusters	0.033	385	2242	1.133	5.729	<0.0001

(B) Pairwise comparisons. Fisher's *p* values in lower triangle, Fisher distances in upper triangle (computed in XLSTAT 2016)

	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5	Cluster 6
Cluster 1	0	36.486	57.562	329.498	149.179	89.374
Cluster 2	**<0.0001**	0	78.831	105.936	83.637	50.733
Cluster 3	**<0.0001**	**<0.0001**	0	521.101	297.414	246.932
Cluster 4	**<0.0001**	**<0.0001**	**<0.0001**	0	47.320	94.767
Cluster 5	**<0.0001**	**<0.0001**	**<0.0001**	**<0.0001**	0	31.044
Cluster 6	**<0.0001**	**<0.0001**	**<0.0001**	**<0.0001**	**<0.0001**	0

(C) Discrimination analysis (computed in XLSTAT 2016)

From\to	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5	Cluster 6	Total	% correct
Cluster 1	19	0	1	0	0	0	20	95.00%
Cluster 2	0	19	0	0	0	1	20	95.00%
Cluster 3	0	0	20	0	0	0	20	100.00%
Cluster 4	0	0	0	19	0	1	20	95.00%
Cluster 5	0	0	0	0	19	1	20	95.00%
Cluster 6	0	0	2	0	0	18	20	90.00%
Total	19	19	23	19	19	21	120	95.00%

Values displayed in bold are significant at the level alpha = 0.003.

related to the third principal component. Accessions with lower eigenvalues at PC3, such as 65351 and 65347, had lower pointness values.

For roundness, the values ranged from 2.1 to 3.4. A roundness of 1.0 denotes a perfect circle, while larger values indicate oblong and non-circular objects. This indicates that the germplasm accessions observed were oblong to non-circular as verified by the descriptor states formulated. Narrowly elliptic descriptor states and those with higher length to width ratio have higher roundness values compared with descriptor states with lower length to width ratio. This is exemplified by accession 66183 having narrowly elliptic kernels with the lowest score for roundness. On the other hand, L4, with the highest roundness value was classified to be obovate. For eccentricity, similar trends can be observed, as a circle will have a value of zero, whereas a straight line will have a value of 1 (Birchfield, 2016). Pointedness, on the other hand, defines how subulate is the basal part of the kernel. Highly pointed accessions will have higher pointness values which are related to the third principal component. Accessions with lower eigenvalues at PC3, such as 65351 and 65347, had lower pointness values. This indicates that the Fourier coefficients can account for the shape variations measured from the calculated kernel shape properties.

Supplementary material

To view supplementary material for this article, please visit https://doi.org/10.1017/S1479262119000285.

Acknowledgements

The authors would like to thank the Department of Science and Technology-Philippine Council for Agriculture, Aquatic and Natural Resources Research and Development (DOST-PCAARRD), Department of Science and Technology-Philippine Council for Health Research and Development (DOST-PCHRD), the T.T. Chang Genetic Resources Center-International Rice Research Institute (IRRI), University of Tsukuba, and the National Plant Genetic Resources Laboratory (NPGRL), College of Agriculture

https://doi.org/10.1017/S1479262119000285 Published online by Cambridge University Press

Food Science, University of the Philippines Los Baños for their support.

References

Bioversity International (2007) *Guidelines for the Development of Crop Descriptor Lists*. Bioversity Technical Bulletin Series. Bioversity International, Rome, Italy.

Birchfield S (2016) Image Processing and Analysis. Cengage Learning. 718 p.

Cervantes E, Martín JJ and Saadaoui E (2016) Updated methods for seed shape analysis. *Scientifica 2016*: 1–10. doi: 10.1155/2016/5691825.

Coronel RE (2011) Important and Underutilized Edible Fruits of the Philippines, University of the Philippines Los Baños Foundation Inc. and Department of Agriculture Bureau of Agricultural Research, Diliman, Quezon City, Philippines.

Freeman H (1974) Computer processing of line drawing images. *Computing Surveys* 6: 57–97.

Gentallan RP, Altoveros NC, Borromeo TH, Endonela LE, Hay FR, Lalusin AG, Reaño CE and Yoshioka Y (2019) Systematic establishment of colour descriptor states through image-based phenotyping. *Plant Genetic Resources* 17: 91–94.

Iwata H and Ukai Y (2002) SHAPE: a computer program package for quantitative evaluation of biological shapes based on elliptic Fourier descriptors. *The Journal of Heredity* 93: 384–385.

Iwata H, Ebana K, Uga Y and Hayashi T (2015) Genomic prediction of biological shape: elliptic Fourier analysis and kernel partial least squares (PLS) regression applied to grain shape prediction in rice (*Oryza sativa* L.). *PLoS ONE* 10: e0120610.

Kuhl FP and Giardina CR (1982) Elliptic Fourier features of a closed contour. *Computer Graphics and Image Processing* 18: 236–258.

Maeda H, Akagi T and Tao R (2018) Quantitative characterization of fruit shape and its differentiation pattern in diverse persimmon (*Diospyros kaki*) cultivars. *Scientia Horticulturae* 228: 41–48.

Neto AJS, Grossi JAS and Pinto FDADC (2009) Postharvest shape characterization of calla lily flowers using digital imaging. *Israel Journal of Plant Sciences* 57: 421–428.

Rohlf FJ (2002) *NTSYSpc: Numerical Taxonomy System*, Version 2.1. Setauket, New York: Exeter Publishing, Ltd.

Rohlf FJ and Archie JW (1984) A comparison of Fourier methods for the description of wing shape in mosquitoes (*Ritera culicidae*). *Systematic Zoology* 33: 322–317.

Sayinci B, Kara M, Ercişli S, Duyar Ö and Ertürk Y (2014) Elliptic Fourier analysis for shape distinction of Turkish hazelnut cultivars. *Erwerbs-Obstbau* 57: 1–11.

Simpson MG (2010) Plant morphology. In: Simpson MG (ed.) *Plant Systematics*, 2nd edn. San Diego: Academic Press, pp. 495–497.

White R, Rentice HC and Verwist T (1988) Automated image acquisition and morphometric description. *Canadian Journal of Botany* 66: 450–459.

Wycislo AP, Clark JR and Karcher DE (2008) Fruit shape analysis of *Vitis* using digital photography. *HortScience* 43: 677–680.

Yoshioka Y, Iwata H, Ohsawa R and Ninomiya S (2004) Analysis of petal shape variation of *Primula sieboldii* by elliptic Fourier descriptors and principal component analysis. *Annals of Botany* 94: 657–664.

https://doi.org/10.1017/S1479262119000285 Published online by Cambridge University Press