# GMR

# Digital phenotyping for quantification of genetic diversity in inbred guava (*Psidium guajava*) families

**W. Krause[1], A.P. Viana[2], N.R. Cavalcante[2], M. Ambrósio[2], E.A. Santos[2] and H.D. Vieira[2]**

[1]Centro de Pesquisa, Estudos e Desenvolvimento Agro-Ambiental,
Universidade do Estado de Mato Grosso, Tangará da Serra, MT, Brasil
[2]Centro de Ciências e Tecnologias Agropecuárias,
Universidade Estadual do Norte Fluminense Darcy Ribeiro,
Campos dos Goytacazes, RJ, Brasil

Corresponding author: W. Krause
E-mail: krause@unemat.br

**ABSTRACT.** Digital image analysis of seeds has been used for the identification of cultivars, determination of seed color and mechanical damage, and classification of different seed sizes. The aim of the present study was to evaluate the efficiency of digital image analysis of seeds for the quantification of genetic diversity among genotypes of inbred guava (*Psidium guajava* L.) families. The SAS Mini equipment, which consists of a capture module and a software program for analysis, was employed for the capture and analysis of the seed images. Different genetic diversity quantification strategies were tested using the Ward-Modified Location Model method. The set of variables related to geometry of the seeds was the largest contributor to divergence among the guava genotypes. The use of seed descriptors obtained by digital image analysis via the SAS system was efficient at quantifying the

genetic diversity among genotypes of inbred guava families associated with the use of the Ward-Modified Location Model method.

**Key words:** Ward-Modified Location Model; Image analysis; Breeding; Digital imaging techniques

## INTRODUCTION

The guava [*Psidium guajava* L. (fam. Myrtaceae)] is a native tropical American tree that is widespread from Mexico to Brazil. It is a species of high economic value for several countries, since many products are generated from it, including fresh fruit, sweets, jellies, juice, syrup, and dried and canned goods. With respect to its nutritional qualities, it is rich in ascorbic acid, calcium, vitamin E, fiber, lycopene, and vitamins A, B6, and B2 (Lima et al., 2002).

The State University of Northern Rio de Janeiro (UENF) has been developing a guava (*P. guajava* L.) breeding program for approximately nine years, generating promising results regarding the composition of several types of study populations (Pessanha et al., 2011; Campos et al., 2013; Oliveira et al., 2013; Campos et al., 2016). In these studies, the production of inbred families was determined as the possibility of forming new types of populations. A cross between related or self-pollinated individuals causes an increase in homozygosis and a decrease in heterozygosis in the offspring, which could be an alternative way to obtain homogeneous fruits in commercial guava orchards since the self-pollination process is an alternative method for the fixation of alleles for the traits of interest (Falconer and Mackay, 1996). Obtaining these types of populations allows recessive alleles for unfavorable effects that are hidden in heterozygotes to manifest in the homozygotes, causing a reduction in the adaptive value of the individual. This is called inbreeding depression or loss of vigor; in the case of guava, this phenomenon has not been observed, which makes possible the use of these strategies in breeding programs for this crop.

To our knowledge, there are no reports of in-depth studies on the effects of inbreeding and its use in *P. guajava* breeding programs, but within the family Myrtaceae, some studies can be found on the genus *Eucalyptus*. Aiming to increase the efficiency of their eucalyptus breeding program, Bison et al. (2004) worked with the effects of inbreeding on the developed populations. Depression varied among the studied clones, averaging 17.5 to 4% for the several variables analyzed. Depression was expected to be higher, but one of the reasons for this outcome is that the self-pollination rate in eucalyptus ranges from 10 to 30%. However, according to the authors, some of the lethal alleles are eliminated naturally. Mortality also occurs during the establishment of seedlings because of the expression of deleterious alleles, and this mortality may be present throughout the different stages of the tree's life cycle. Because it was possible to select plants in the $F_2$ generation with performance superior to that of the $F_1$ parents in most clones, the authors concluded that clones originating from self-pollination are a good alternative for obtaining superior plants, since plants with lethal alleles can be eliminated in advance during the seedling production stage.

In a breeding program, the breeder needs variability in the population, and not only at the beginning of the program. Rather, they must also monitor whether, over its course, the population is losing variability, which may cause future genetic gains to become unavailable and prevent the continuity of the program. This knowledge allows the breeder to perform the introgression of favorable alleles for traits of interest at any time in the breeding program when he or she realizes that the genetic variability in the population is small or agronomic

performance is low. To measure genetic variability, morphological descriptors of several parts of the plant can be used, such as those related to the leaf, flower, fruit, seeds, and DNA markers. In the case of guava, many variables and methodologies have been adopted successfully to determine genetic diversity among genotypes (Pessanha et al., 2011; Campos et al., 2013; Oliveira et al., 2013; Campos et al., 2016).

The seed is the fertilized and developed ovum of superior plants, and its importance is related to reproduction and dispersion. Therefore, the phenotyping of this organ is of extreme relevance in breeding programs, given that it is possible to determine genetic diversity for seed-related variables (Gustin and Settles, 2015).

Phenotyping seeds via visual analysis is a laborious process that yields subjective results, further complicating the analysis of seeds for both commercial and technological purposes. However, today new technologies for controlled environments such as spectroscopy, thermography, tomography, fluorescence, isotope discrimination, and digital imaging techniques are being employed to increase the accuracy of and reduce the time required for phenotypic measurements for better characterization (Santos and Yassitepe, 2014). As a result, traits that have never been measured before or that have only been measured in specific situations are starting to be measured more frequently (Fiorani and Schurr, 2013) and with greater ease, e.g., traits related to seed size, color, form, and texture (Granitto et al., 2005). Venora et al. (2007) also reported that image analysis is a fast method that requires less than one minute for digitalization and measurement. It is also a very easily reproducible technique, in addition to being reliable and non-destructive. Images can be acquired by classic reflectance techniques such as light and microscopy photography or by more modern techniques like thermography, fluorescence, tomography, magnetic resonance, and nuclear resonance (Montes et al., 2007; Berger et al., 2010; Houle et al., 2010).

In Brazil, the Seed Analysis System (SAS) equipment was developed by Tbit in 2011 to capture images, histograms, graphs, and other items that facilitate the analysis of seeds and seedlings. The equipment consists of a capture module and analysis software. The object of study is placed on an acrylic tray in capture mode and its image is captured by one or two cameras, depending on the version used. SAS performs an individual analysis of the images of seeds and plantlets, providing a large amount of information on the color, texture, geometry, uniformity, vigor, and growth of seedlings (Andrade et al., 2016).

The digital image analysis of seeds has been employed for the identification of cultivars, determination of seed colors and mechanical damage, and classification of different seed sizes (Venora et al., 2009; Medina et al., 2010; Kara et al., 2013; Pinto et al., 2015; Andrade et al., 2016). However, to our knowledge, there are no reports on the use of digital imaging to measure seed traits such as size, form, texture, and color as applied for the quantification of intraspecific genetic diversity. Thus, the objective of the present study was to evaluate the use of digital image analysis of seeds, obtained via SAS, for the quantification of genetic diversity among genotypes from inbred guava families obtained in a program under development at UENF associated with the use of the Ward-MLM (Modified Location Model) method.

## MATERIAL AND METHODS

### Obtaining the seeds

The experiment was performed at the experimental unit of UENF in Itaocara - Ilha

Barra do Pomba (21°40'S, 42º04'W, 76 m in altitude), in the northwest region of Rio de Janeiro State, Brazil. The soil of the experimental area is classified as a Red-Yellow Argisol.

The experiment was set up as a randomized block design with 18 inbred guava (*P. guajava*) families, 3 replications, and 10 plants per plot. Families were obtained by self-crosses of progenies resulting from controlled biparental crosses (Pessanha et al., 2011) from the breeding program of the State University of Northern Rio de Janeiro (UENF). Planting took place in July 2014, with rows spaced 4 m apart and plants 1.5 m apart. Liming and fertilization at planting and topdressing were performed according to the soil analysis, following the recommendations of Costa and Costa (2003).

In this experiment, 61 genotypes (individual plants) of guava were evaluated. Fifty seeds were collected from five ripe fruits per genotype in the period from March to April 2016. Seeds were removed manually, washed in running water over a fine mesh sieve, and then left to dry for three days on germination paper at room temperature.

## Capture and analysis of guava seed images

SAS mini equipment was used to capture and analyze the seeds. The equipment comprises a capture module and a software program for analysis. The capture module consists of an acrylic tray in which 50 seeds were placed for image capture by the high-resolution camera (Figure 1). The software generates graphs and spreadsheets from the captured information to facilitate interpretation of the seed images. The CIELab color system was used to calibrate the background color, in which lightness ranged from 0.0 to 74.0, component a* from -120.0 to 120.0, and component b* from -120.0 to 0.0. The recognition parameter of the studied object was the seeds in general, and the minimum size was fixed at 0.2 mm (Andrade et al., 2016).
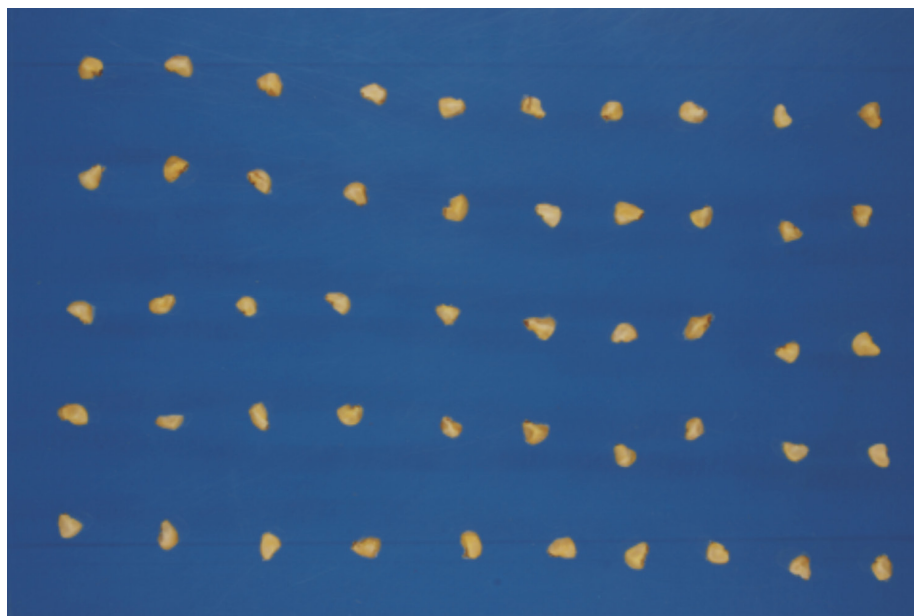


**Figure 1.** Capture module of images from the SAS mini equipment.

SAS generated a spreadsheet with mean values of each genotype for the color, geometry, and texture variables. In this study, 136 seed variables were obtained; 49 of these were related to color, 44 to geometry, and 43 to texture.

## Quantification of genetic divergence

The Ward-MLM (Modified Location Model) method proposed by Franco et al. (1998) was employed as described by Viana and Resende (2014). First, the Gower (1971) index was used, generating an estimate of the dissimilarity index, which ranges from 0 to 1. Dissimilarity was calculated as:

$$Sij = \frac{\sum_{k=1}^{p} W_{ijk}.S_{ijk}}{\sum_{K=1}^{p} W_{ijk}} \qquad \text{(Equation 1)}$$

where i and j = individuals to be compared for trait k; p = total number of traits; and $S_{ij}$ = contribution of variable k to the total distance.

If the variable was qualitative, $S_{ijk}$ assumed value 1 when there was positive or negative agreement for the trait k between individuals i and j; otherwise, when the variable was quantitative,

$$S_{ij} = \frac{\left| Y_{ik} - Y_{jk} \right|}{R_k} \qquad \text{(Equation 2)}$$

where $R_k$ = amplitude of variation of variable k, assuming values of 0 to 1 or integers between them.

The $W_{ijk}$ value was a weight used to define the contributions of the individual $S_{ijk}$'s. When the value of variable k was absent in one or both individuals, $W_{ijk} = 0$; otherwise, it was equal to 1.

Next, the ideal number of groups was defined according to the pseudo F and pseudo $t^2$ criteria by Ward's clustering method (Ward, 1963). According to the optimal number of groups, the hierarchical classification was obtained by Ward's method, which provides the necessary parameters to implement the final step of the MLM model (Crossa and Franco, 2004). The differences between groups, the correlation between variables, and the canonical variable (CV) were analyzed graphically. All analyses were performed using SAS statistical software (SAS Institute, 2009). The diagrams were performed using the Sigma Plot software version 11.0.

For the analyses, different strategies were used to quantify the genetic diversity. First, the WARD-MLM methodology was applied for each group of color, geometry, and texture variables in which the contribution of variables was evaluated to quantify the existing genetic diversity among the genotypes; i.e., which showed the highest correlations with the first canonical variable. Later, three strategies were applied for the quantification of genetic diversity. In the first, the 30% of variables that contributed most to divergence within each

group were used, including 9 color, 12 geometry, and 6 texture variables. In the second strategy, 20% were used (6 color, 8 geometry, and 4 texture variables), and in the third, 10% were used (3 color, 4 geometry, and 2 texture variables).

## RESULTS

Of the 136 evaluated variables, 44 were discarded because there was no variation. Thus, 31 color, 40 geometry, and 21 texture variables were analyzed.

According to the pseudo F and pseudo $t^2$ criteria, three groups were formed for all of the seed descriptor evaluation strategies obtained via digital image analysis using SAS (Table 1). The ideal number of groups was found where there was the highest increase in the logarithmic function. Thus, there was no reduction in the number of groups even after the number of descriptors was reduced to nine.

However, the number of genotypes per group varied between the evaluated strategies (Table 1). In the strategy in which only the 31 color descriptors were evaluated, groups I, II, and III were formed by 43, 12, and 6 genotypes, respectively. For the set of seed-geometry variables, group I was formed by 19, group II by 20, and group III by 22 genotypes. In the analysis of texture variables, group I included 21 genotypes; II, 31; and III, 9. In the evaluation of the strategy in which the 30% of the total descriptors that contributed most to divergence were used, group I was formed by 18 genotypes, II by 37, and III by 6. For the strategy that utilized 20% of the total descriptors, group I had 42 genotypes; II, 12; and III, 7. Lastly, when 10% of the total descriptors were used, group I was formed by 35; II by 4; and III by 22 genotypes (Table 1).

Regarding the distance between the groups formed by the Ward-MLM method for the different seed descriptor evaluation strategies, there was also an alteration in the values. The largest distance was found between groups I and III (218.25) for the 31 color descriptors, and between groups II and III for the geometry (69.37) and texture (65.19) descriptors. In the evaluation of 30% of the total descriptors, the largest distance was between groups II and III (79.46). For the evaluation of 20 and 10% of the total descriptors, however, the largest distance was between groups I and II: 40.09 and 22.41, respectively (Table 1).

**Table 1.** Number of groups, number of genotypes per group, and distance between groups formed by the Ward-MLM method for the different strategies of evaluation of seed descriptors in 61 genotypes of inbred guava families.

| Evaluation strategy | Number of groups | Number of genotypes per group | Distance between groups |
|---|---|---|---|
| 31 color descriptors | I | 43 | I x II = 25.44 |
| | II | 12 | I x III = 218.25 |
| | III | 6 | II x III = 118.93 |
| 40 geometry descriptors | I | 19 | I x II = 34.35 |
| | II | 20 | I x III = 36.37 |
| | III | 22 | II x III = 69.37 |
| 21 texture descriptors | I | 21 | I x II = 20.01 |
| | II | 31 | I x III = 25.09 |
| | III | 9 | II x III = 65.19 |
| 30% of the total descriptors (color: 9; geometry: 12; texture: 6) | I | 18 | I x II = 28.97 |
| | II | 37 | I x III = 38.02 |
| | III | 6 | II x III = 79.46 |
| 20% of the total descriptors (color: 6; geometry: 8; texture: 4) | I | 42 | I x II = 40.09 |
| | II | 12 | I x III = 24.13 |
| | III | 7 | II x III = 32.38 |
| 10% of the total descriptors (color: 3; geometry: 4; texture: 2) | I | 35 | I x II = 22.41 |
| | II | 4 | I x III = 14.13 |
| | III | 22 | II x III = 21.81 |

　　　　The distances between groups were also observed in the graphic analysis (Figure 2). In the analysis based on the canonical variables, the first two canonical variables (CV) obtained by the Ward-MLM method explained 100% of the total variation for all evaluation strategies adopted (Figure 2). This high value indicates that a two-dimensional graph is appropriate for visualizing the relationship between the groups and the genotypes within the groups.
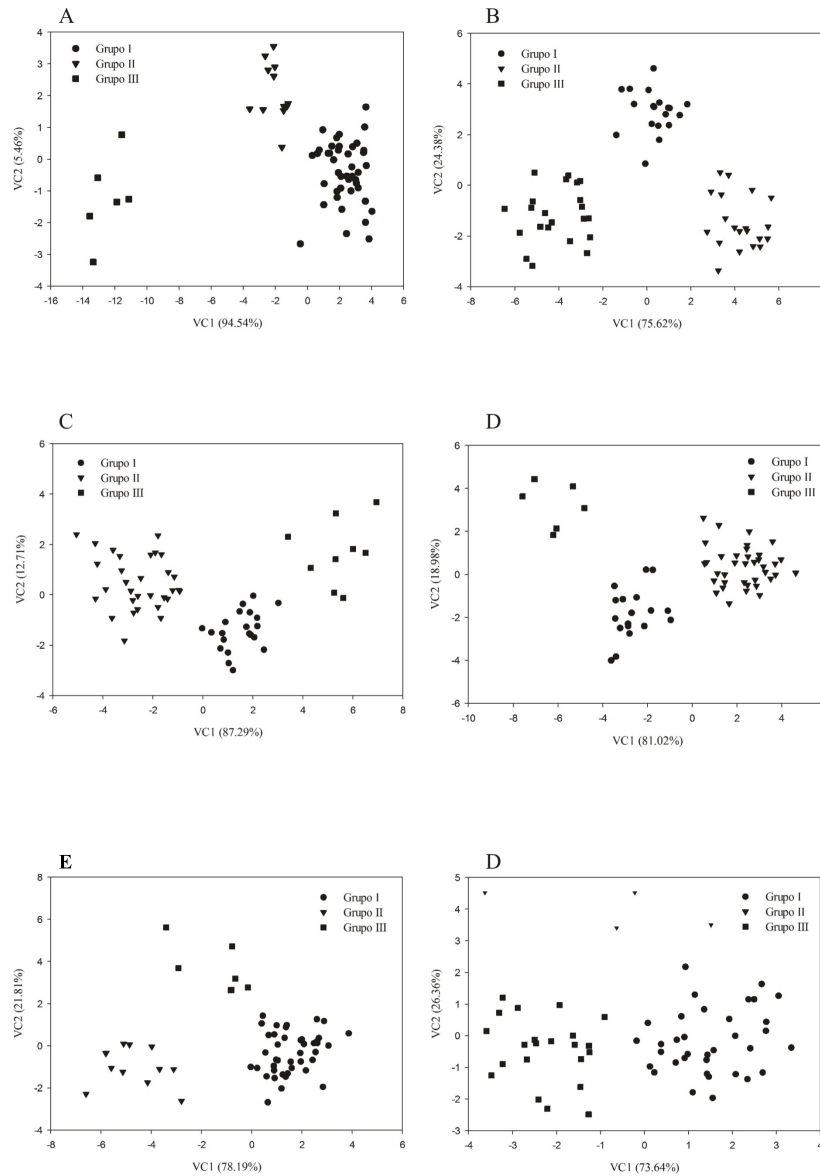


**Figure 2.** Dispersion of the first two canonical variables (CV1 and CV2) representing the formation of three groups by the Ward-MLM strategy based on: **A.** 31 color descriptors; **B.** 40 geometry descriptors; **C.** 21 texture descriptors; **D.** 30% of the total seed descriptors; **E.** 20% of the total descriptors; and **F.** 10% of the total descriptors.

Table 2 shows the values of the contribution of variables to quantifying the existing genetic diversity between the genotypes; i.e., which ones presented the highest correlations with the first canonical variables for the strategies evaluating 30, 20, and 10% of the descriptors. In the three strategies, the set of variables regarding geometry were the largest contributors to quantifying the existing genetic diversity among genotypes, followed by the texture variables and lastly the color-related variables.

**Table 2.** Canonical variables for the different strategies of evaluation of seed descriptors in 61 genotypes of inbred guava families.

| Descriptor | | Evaluation strategy[1/] | | | | | |
|---|---|---|---|---|---|---|---|
| | | 30% | | 20% | | 10% | |
| | | CV1 | CV2 | CV1 | CV2 | CV1 | CV2 |
| Color | Below Otsu: blue band | 0.060 | 0.496 | 0.337 | 0.406 | - | - |
| | Below Otsu: green band | 0.136 | 0.361 | 0.339 | 0.248 | 0.221 | 0.201 |
| | Below Otsu: CIELab: a | -0.011 | -0.307 | -0.191 | -0.200 | - | - |
| | Below Otsu: CIELab: L | 0.152 | 0.356 | 0.353 | 0.242 | 0.234 | 0.186 |
| | CIELab: dispersion | 0.029 | -0.518 | - | - | - | - |
| | CIELab: a dispersion | 0.015 | -0.451 | -0.325 | -0.435 | -0.148 | -0.326 |
| | CIELab: L | 0.064 | 0.342 | - | - | - | - |
| | Intensity | 0.106 | 0.406 | 0.310 | 0.321 | - | - |
| | Lightness | 0.106 | 0.395 | - | - | - | - |
| | Circularity by FFCg | 0.775 | -0.046 | - | - | - | - |
| Geometry | Feret diameter | -0.814 | 0.162 | -0.571 | 0.678 | - | - |
| | Maximum diameter (cm) | -0.778 | 0.184 | - | - | - | - |
| | Circumference (cm) | -0.813 | 0.130 | -0.564 | 0.633 | -0.679 | 0.474 |
| | Convex circumference (cm) | -0.815 | 0.164 | -0.568 | 0.660 | -0.676 | 0.530 |
| | Rectangles: largest area (cm$^2$) | -0.822 | 0.182 | -0.533 | 0.711 | - | - |
| | Rectangles: largest edge (cm) | -0.819 | 0.173 | - | - | - | - |
| | Rectangles: largest circumference (cm) | -0.815 | 0.171 | -0.560 | 0.656 | -0.661 | 0.525 |
| | Rectangles: average largest edges (cm) | -0.807 | 0.166 | - | - | - | - |
| | Rectangles: average smallest edges (cm) | -0.792 | 0.192 | -0.548 | 0.660 | - | - |
| | Rectangles: smallest circumference (cm) | -0.807 | 0.180 | -0.554 | 0.676 | - | - |
| | Rectangles: average circumference (cm) | -0.811 | 0.178 | -0.558 | 0.671 | -0.652 | 0.554 |
| Texture | Laws: ER | -0.767 | 0.256 | -0.444 | 0.723 | | |
| | Laws: EW | -0.760 | 0.256 | -0.447 | 0.715 | -0.599 | 0.497 |
| | Laws: LW | -0.763 | 0.294 | - | - | - | - |
| | Laws: RW | -0.757 | 0.269 | - | - | - | - |
| | Laws: SR | -0.761 | 0.275 | -0.427 | 0.735 | - | - |
| | Laws: SW | -0.755 | 0.272 | -0.427 | 0.732 | -0.574 | 0.519 |

[1/]30% of total descriptors (color: 9; descriptors: 12; texture: 6); 20% of total descriptors (color: 6; geometry: 8; texture: 4); 10% of total descriptors (color: 3; geometry: 4; texture: 2).

## DISCUSSION

The use of likelihood function analysis can define more precise criteria in the formation of groups, resulting in the determination of less-subjective groups (Barbé et al., 2010). Campos et al. (2013) quantified the divergence among 138 guava genotypes obtained from controlled biparental crosses and found eight groups, with an increase of 67.51. However, the number of groups may vary according to species, number of accessions, and number and type of descriptor (Gonçalves et al., 2009). Another fact that explains the formation of a lower number of groups in this study is that the genotypes originated from inbred guava families.

The formation of three groups also showed that the use of a set of variables referring to color, geometry, and texture measured by digital analysis of seed images obtained via SAS was efficient at quantifying the genetic diversity among the genotypes of the inbred guava families evaluated here, even when only nine descriptors were used (Table 1). In this way,

these variables are useful as additional descriptors in the evaluation of genetic variability.

Tourian and Padilha (2008) aimed to identify eight varieties of soybean according to the shape and source of the seeds using digital image processing aided by Artificial Neural Networks (ANN). Analysis of the morphological properties of the soy seeds by digital image processing together with RNA enabled the distinction between soy cultivars.

With respect to the analysis of canonical variables in this study, the first two variables explained 100% of the variability for all evaluated strategies (Figure 2). Thus, the two-dimensional graphic representation was a satisfactory interpretation of the variability among the genotypes, because the first two canonical variables allowed estimates of greater than 80% of the total variation (Cruz et al., 2012).

Campos et al. (2013) evaluated 138 guava accessions based on morphological, agronomic, and physicochemical descriptors, and found that the first two variables were responsible for only 61.79% of the variability. Canonical variable 1 was responsible for 39.12% and CV2 for 22.68% of the total variation. Thus, a two-dimensional graph would not capture a good proportion of total variability, thereby requiring the introduction of the third variable, which accounted for 19.50% of the variation. Therefore, the tridimensional representation was the most suitable for the representation of the data set, and the sum of the three variables accounted for 81.30% of the total variation.

Kara et al. (2013) utilized seed size and shape properties, which were determined using the image processing method in digital photography to group 12 varieties of beans in Turkey. According to the results of the principal components, the most important variables of the first component were weight, average geometric diameter, surface area, volume, projected area, equivalent diameter, perimeter, and length. The authors also found that the first two components explained 96.1% of the total variation, with four groups formed.

Principal component analysis was used by Medina et al. (2010) in the investigation of a set of 38 variables composed of 8 geometrical, 20 morphological, four fractal, and six color variables obtained from 25 varieties of quinoa (*Chenopodium quinoa* Willd) seeds. The first principal component (PC1) explained 47.7% of the total variation; the second plus the first, 65.8%; and the sum of the first three components explained 75.5%, suggesting that a three-dimensional graphic representation would be more suitable. The authors also observed that the European varieties were grouped separately by both the principal component analysis and UPGMA. Thus, the results demonstrated the possibility of separating the 25 quinoa seed varieties by geographic origin.

Analysis of the contribution of the variables to quantifying the existing genetic diversity among the genotypes showed that the set of geometry-related variables was the largest contributor to quantifying the existing genetic diversity among the genotypes, followed by the sets of texture and color variables (Table 2). Because of the small size and irregular shape of the guava seeds, obtaining the set of geometry variables by visual analysis would be a laborious process with subjective results. However, with the analysis of digital images obtained via SAS, measurement is rapid and efficient, facilitating the use of these variables in the quantification of genetic variability.

Venora et al. (2007) worked with lentil seeds and used a digital image analysis method developed in Canada in which the image is captured by a flatbed scanner. The authors used the equipment to measure the size, shape, and color of the seeds and differentiate five varieties of lentil. They mentioned that the method typically used to evaluate these traits is visual analysis of the seed, which is slow and subjective. The digital image analysis was faster, and the time

spent on digitalization and measurement was less than one minute. Furthermore, the authors stressed that image analysis is a safe, non-destructive method with high repeatability.

In Italy, 15 crossbred bean varieties were evaluated using seed shape, size, color, and texture characteristics obtained via digital analysis to identify the growing region of these varieties (Venora et al., 2009). The authors demonstrated that it was possible to differentiate the bean varieties according to their growing region, thus enabling their traceability, and consequently perform the geographical indication. Moreover, they concluded that this method of digital image analysis of seeds is rapid, requiring less than one minute for digitalization and measurement, in addition to being easily repeatable, reliable, and non-destructive. It is also a low-cost identification method, as it does not require any type of chemical reagent.

## CONCLUSIONS

The set of variables related to seed geometry was the largest contributor to divergence among the guava genotypes.

The adoption of seed descriptors obtained via digital image analysis using the SAS system was efficient at the quantification of genetic divergence among genotypes of inbred guava families associated with the Ward-MLM (Modified Location Model) method.

### Conflicts of interest

The authors declare no conflict of interest.

## ACKNOWLEDGMENTS

## REFERENCES

Andrade DB, Oliveira A dos S, Pinto CAG, Pires RM de O, et al. (2016). Detection of green seeds in soybean lots by the seed analysis system (SAS). *Int. J. Curr. Res.* 8: 26,462-26,465.

Barbé TC, Amaral Júnior AT, Gonçalves LSA, Rodrigues R, et al. (2010). Association between advanced generations and genealogy in inbred lines of snap bean by the Ward-Modified Location Model. *Euphytica* 173: 337-343. http://dx.doi.org/10.1007/s10681-009-0089-z

Berger B, Parent B and Tester M (2010). High-throughput shoot imaging to study drought responses. *J. Exp. Bot.* 61: 3519-3528. PubMed http://dx.doi.org/10.1093/jxb/erq201

Bison O, Aguiar AM, Resende GDSP and Ramalho MAP (2004). Inbreeding depression in Eucalyptus clones. *Crop Breed. Appl. Biotechnol.* 4: 459-464. http://dx.doi.org/10.12702/1984-7033.v04n04a13

Campos BM, Viana AP, Quintal SSR, Gonçalves LSA, et al. (2013). Quantification of the genetic divergence among guava accessions using Ward-MLM strategy. *Rev. Bras. Frutic.* 35: 571-578. http://dx.doi.org/10.1590/S0100-29452013000200028

Campos BM, Viana AP, Quintal SSR, Barbosa CD, et al. (2016). Heterotic group formation in *Psidium guajava* L. by artificial neural network and discriminant analysis. *Rev. Bras. Frutic.* 38: 151-157. http://dx.doi.org/10.1590/0100-2945-258/14

Costa AFS and Costa AN (2003). Tecnologias para produção de goiaba. Instituto Capixaba de Pesquisa e extensão Rural, Vitória.

Crossa J and Franco J (2004). Statistical methods for classifying genotypes. *Euphytica* 137: 19-37. http://dx.doi.org/10.1023/B:EUPH.0000040500.86428.e8

Cruz CD, Regazzi AJ and Carneiro PCS (2012). Modelos biométricos aplicados ao melhoramento genético. Editora UFV, Viçosa.

Falconer DS and Mackay TFC (1996). Introduction to quantitative genetics. 4th ed. Longman Group Limited, Edinburgh.

Fiorani F and Schurr U (2013). Future scenarios for plant phenotyping. *Annu. Rev. Plant Biol.* 64: 267-291. PubMed http://dx.doi.org/10.1146/annurev-arplant-050312-120137

Franco J, Crossa J, Villasenõr J and Taba S (1998). Classifying genetic resources by categorical and continuous variables. *Crop Sci.* 38: 1688-1696.

Gonçalves LS, Rodrigues R, do Amaral Júnior AT, Karasawa M, et al. (2009). Heirloom tomato gene bank: assessing genetic divergence based on morphological, agronomic and molecular data using a Ward-modified location model. *Genet. Mol. Res.* 8: 364-374. PubMed http://dx.doi.org/10.4238/vol8-1gmr549

Gower JC (1971). A general coefficient of similarity and some of its properties. *Biometrics* 27: 623-637. http://dx.doi.org/10.2307/2528823

Granitto PM, Verdes PF and Ceccato HA (2005). Large-scale investigation of weed seed identification by machine vision. *Comput. Electron. Agric.* 47: 15-24. http://dx.doi.org/10.1016/j.compag.2004.10.003

Gustin JL and Settles AM (2015). Fenômica de sementes. In: Fenômica: como a fenotipagem da próxima geração está revolucionando o melhoramento de plantas (Fritsche-Neto R and Borém A, eds.). Editora UFV, Viçosa, 103-126.

Houle D, Govindaraju DR and Omholt S (2010). Phenomics: the next challenge. *Nat. Rev. Genet.* 11: 855-866. PubMed http://dx.doi.org/10.1038/nrg2897

Kara M, Sayinci B, Elkoca E, Öztürk İ, et al. (2013). Seed size and shape analysis of registered common bean (*Phaseolus vulgaris* L.) cultivars in Turkey using digital photography. *J. Agric. Sci.* 19: 219-234.

Lima MAC, Assis JS and Neto LG (2002). Caracterização dos frutos de goiabeiras e seleção de cultivares na região do sub-médio São Francisco. *Rev. Bras. Frutic.* 24: 273-276. http://dx.doi.org/10.1590/S0100-29452002000100061

Medina W, Skurtysb O and Aguilera JM (2010). Study on image analysis application for identification Quinoa seeds (Chenopodium quinoa Willd) geographical provenance. *Food Sci. Technol.* 43: 238-246.

Montes JM, Melchinger AE and Reif JC (2007). Novel throughput phenotyping platforms in plant genetic studies. *Trends Plant Sci.* 12: 433-436. PubMed http://dx.doi.org/10.1016/j.tplants.2007.08.006

Oliveira NNS, Viana AP, Quintal SSR, Paiva CL, et al. (2013). Análise de distância genética entre acessos do gênero *Psidium* via marcadores ISSR. *Rev. Bras. Frutic.* 36: 917-923. http://dx.doi.org/10.1590/0100-2945-413/13

Pessanha PGO, Viana AP, Amaral Júnior AT, Souza RM, et al. (2011). Avaliação da Diversidade Genética em Acessos de *Psidum ssp.* via marcadores RAPD. *Rev. Bras. Frutic.* 33: 129-136. http://dx.doi.org/10.1590/S0100-29452011000100018

Pinto CAG, de Carvalho MLM, de Andrade DB, Leite ER, et al. (2015). Image analysis in the evaluation of the physiological potential of maize seeds. *Rev. Cienc. Agron.* 46: 319-328.

Santos TT and Yassitepe JE de CT (2014). Fenotipagem de plantas em larga escala: um novo campo de aplicação para a visão computacional na agricultura. In: Tecnologias da Informação e Comunicação e suas relações com a agricultura (Massruhá SMFS, Leite MA de A, Luchiari Junior A and Romani LAS, eds.). Embrapa Informática Agropecuária, Campinas, 85-100.

SAS Institute (2009). Statistical Analysis System: User's Guide. SAS, Cary.

Tourian OK and Padilha FRR (2008). Reconhecimento de variedades de soja por meio do processamento de imagens digitais usando redes neurais artificiais. *Eng. Agric.* 28: 759-769. http://dx.doi.org/10.1590/S0100-69162008000400016

Venora G, Grillo O, Shahin MA and Symons SJ (2007). Identification of Sicilian landraces and Canadian cultivars of lentil using an image analysis system. *Food Res. Int.* 40: 161-166. http://dx.doi.org/10.1016/j.foodres.2006.09.001

Venora G, Grillo O, Ravalli C and Cremonini R (2009). Identification of Italian landraces of bean (Phaseolus vulgaris L.) using an image analysis system. *Sci. Hortic. (Amsterdam)* 121: 410-418. http://dx.doi.org/10.1016/j.scienta.2009.03.014

Viana AP and Resende MDV (2014). Genética quantitativa aplicada no melhoramento de fruteiras. Editora Interciência, Viçosa.

Ward JH (1963). Hierarchical grouping to optimize an objective function. *J. Am. Stat. Assoc.* 58: 236-244. http://dx.doi.org/10.1080/01621459.1963.10500845