Image-based phenotyping of morpho-agronomic traits in papaya fruits (Carica papaya L. THB var.)

Renato Santa-Catarina*1, Diego Fernando Marmolejo Cortes1, Julio Cesar Fiorio Vettorazzi1, Helaine Christine Cancela Ramos1, Geraldo Antônio Ferreguetti2 and Messias Gonzaga Pereira1

1Department of Genetic and Plant Breeding, North Fluminense State University “Darcy Ribeiro”, Av. Alberto Lamego, 2000, 28013-602, Campos dos Goytacazes, Rio de Janeiro, Brazil
2Caliman Agrícola S.A., BR 101, km 111, Linhares, Espírito Santo, Brazil

*Corresponding author: renato.scat@gmail.com

Abstract

Manual phenotyping (MP) methodology for papaya breeding demands intensive labor and is time-consuming. This study aimed to validate a low-cost image-based phenotyping (IBP) methodology of fruit traits to speed up the process in breeding program. Two groups of 50 fruits of the ‘THB’ at the zero-maturity stage were used. The fruits were sliced longitudinally and half part of the fruit was scanned in Flat-bed Scanner to obtain the digital image. In the first group, the length and diameter of fruits were measured by image processing (IP) using a two-bit binary image and the ovarian cavity length, ovarian cavity diameter, thickness pulp were measured by image analysis (IA) in RGB format using the straight tool of ImageJ software. The traits were measured by digital caliper. The fruit mass (FM) was measured by analytical scale and de fruit volume (FV) was estimated using the water displacement method (WDM). The second group was used to estimate the FV and FM through IP and MP. The trait averages were estimated using IBP, similar to those averages measured manually. The coefficients of variation estimated by IBP were low compared to the measurement by MP, indicating that this methodology is more accurate. The Bland-Altman approach showed agreement between the FV obtained by the WDM and IP. Since the IBP is about four times quicker and less-dependent on labor, it is expected to be incorporated to papaya breeding as a way to increase the number of accessed (being evaluated) genotypes and, consequently, increasing the genetic gains.

Keywords: Brazil; Carica papaya L.; Digital Phenotyping; Fruit Volume; Plant Breeding.

Abbreviations: UENF_State University of North Fluminense “Darcy Ribeiro”; MP_Manual phenotyping; IBP_Image-based phenotyping; IA_Image analysis; IP_Image processing; CV_coefficients of variation; WDM_water displaced method; CI_Confidence interval; TP_Thickness pulp; OCL_Ovarian cavity length; OCD_Ovarian cavity diameter; Fl_Fruit length; FD_Fruit diameter; FV_Fruit volume; FM_Fruit mass.

Introduction

Brazil stands out as one of the world’s leading producers of papaya Carica papaya (L.) being India the first (FAO, 2018). Papaya is one of the most cultivated and consumed fruit crops in tropical and subtropical regions of the world. The fruits are sources of vitamins, minerals and proteins. They also contain the alkaloid carpaine and the papain enzyme used in medicine, textile industry and cooking (Souza et al., 2005).

The evaluation of fruits for commercial purposes and breeding programs is performed manually in laboratory (Marin et al., 2006; Silva et al., 2007; Silva et al., 2008, Barros et al., 2017). This activity requires an intensive labor which limits the evaluation of a large number of individuals and may decrease experimental precision, reduce the chances of selecting superior genotypes as well as obtaining genetic gain (Rosher et al., 2014).

In order to improve the efficiency of phenotyping for traits of interest, some methodologies of image analysis and processing were developed to allow high performance phenotyping, increasing the number of assessed genotypes, reducing time and experimental error (Li et al., 2014). In agriculture, application of image processing embraces several areas, from estimates of levels of fertilization and irrigation to the prediction of yield and post-harvest for the determination of fruit quality and size (Sablov et al., 2002). Recently, image-based phenotyping has been used to inspect agricultural products automating the evaluation process with accuracy and practicality, aerial monitoring of farms for determination and recognition of diseases and abiotic stress and estimation of morphological traits (Rashidi et al., 2007; Pandey et al., 2013). The image-based phenotyping uses image processing and analysis techniques. These techniques are computer vision part, which is considered a science that studies the theoretical and algorithmic foundations for extracting useful information from an object through images (Gunasekaran, 2000). This science implementation requires 2D or 3D images and includes the use of systems based on visible light, infrared, laser, X-ray and magnetic systems (Pandey et al., 2013).
The image-based phenotyping has been validated with high degree of accuracy to estimate the length, diameter and volume in watermelon (Koc, 2007), kiwi and melon (Rashidi et al., 2007; Rashidi et al., 2009), in tangerine (Khojastehnazhand et al., 2008), cantaloupe fruits (Rashidi et al., 2009), mango (Teoh et al., 2014), lemon (Khojastehnazhand et al., 2010) and citrus fruits (Omid et al., 2010). In papaya breeding evaluations, image-based phenotyping is a great option for measuring fruit traits to evaluate a larger group of genotypes with speed and accuracy.

In papaya breeding, the fruit mass is an important yield component as well as a qualitative aspect to cultivar development and market acceptance. Fruit mass is also related with the transport and commercialization (Sabolov et al., 2002; Koc, 2007).

This study aims to validate a low-cost methodology of image-based phenotyping that can be used in papaya breeding programs to genetic increment gains during the selection cycles.

Results and discussion

Manual phenotyping (MP) x Image-based phenotyping (IBP)

Table 1 shows the statistics obtained from the analysis of both manual phenotyping and image-based phenotyping of the morpho-agronomic fruit traits measured in the ‘THB’ variety. For the two methodologies, similar averages were obtained for fruit length (FL), fruit diameter (FD) ovarian cavity length (OCL) ovarian cavity diameter (OCD), average the thickness Pulp (TP), fruit volume (FV), and fruit mass (FM). The paired t-test showed no differences (p > 0.05) between the averages for all traits estimated through IBP methodology and through MP. The estimate error using digital images ranged from 0.15 % for the FL to 0.70 % for the OCD traits, showing that this IBP method can be deployed to estimate these traits with a high degree of accuracy. Several authors have reported success in estimating the diameter and length of fruits using digital images. For instance, Koc (2007) validated the FL and FD traits in watermelon with the use of digital images, obtaining averages similar to the measurements taken with a caliper. Similarly, Rashidi et al. (2009), in cantaloupe and Khojastehnazhand et al. (2009) in orange found that the fruit averages obtained with digital images were similar to those measured with a caliper, confirming that IBP methodology is efficient for phenotyping of fruit traits. Other authors find similar results using IP methodology to estimate the length, diameter and fruit volume (Rashidi et al., 2007; Teoh et al., 2014; Wang et al., 2007).

It could be seen that the correlations between the manual measurements and those using digital image analysis for the all morphological fruit traits were high, ranging from 0.94 for fruit diameter (FD) to 0.99 for fruit length (FL) and ovarian cavity length (OCL) (Table 2), showing that the IBP phenotyping methodology presented good constancy. Coefficient of variation (CV) is used to express the variability of statistical data excluding the influence of magnitude of the variable. The lower CV value, the more homogeneous the data will be, the smaller the dispersion around the average and the greater the experimental precision. The variation coefficients obtained by the IBP methodology and the MP for all traits were lowest. It is interesting to note that the CV was lower for the IBP methodology to all traits, indicating that the IBP methodology has more precision, when compared to the MP methodology (Table 1).

These results show that the phenotyping of morphological papaya fruit traits by the IBP methodology enables accurate estimations and can be used to replace the traditional methodology in the papaya breeding program.

Estimated fruit volume

Since there was no difference between the MP methodology and the IBP measurements, it was possible to estimate the volume using IP methodology by the equation 1. The paired samples t-test results show that the FV was determined through IP methodology is not different (P > 0.05) from the FV measured via water displacement methodology (WDM). Fruit volume average determined through the WDM was 473.50 mL and the average FV determined through the IP was 480.32 mL. The CV was lower for the IP methodology to FV compared to the WDM, indicating that the IP methodology has more precision when compared to the WDM (Table 2).

The correlation between the volume calculated by WDM and by the IP using the dimensions obtained with the digital image was high (r = 0.95) (Figure 1a). According to the Bland and Altman (1999) approach, these methods are consistent, which means that the volume of papaya fruits may be estimated using equation 1. The average difference between the two methods is -7 mL.

The FV differences between IP and WDM methodologies were normally distributed and 95% of the FV differences were expected to lie between μ–1.96α and μ+ 1.96α, known as 95% limits of agreement (Bland and Altman, 1999). The 95% limits of agreement of the differences between the volume calculated by WDM and IP were -62.88 and 49.24 mL. The volumes by expression of ellipsoid approximation can be lower than 62.88 mL or 49.24 mL higher than the volumes calculated by WDM (Figure 1b).

The estimate error using digital images was 1.42 % for the FV trait, showing that IBP method can be deployed to estimate this trait with high degree of accuracy. The errors in this methodology may be due to the use of an approach expression, which is based on the assumption that the fruits have a perfectly ellipsoid shape. Therefore, irregularly shaped fruits serve to increase the errors of this methodology. The Bland and Altman (1999) approach has been used to validate the volume and area of fruits using IBP. Koc (2007), Khojastehnazhand et al. (2008), Khojastehnazhand et al. (2009), Rashidi et al. (2009) and Arjenaki et al. (2012) have used this approach for estimation of watermelon, tangerine, orange, cantaloupe, and the surface area of zucchini, respectively.

Estimated fruit mass

Fruit mass is a very important trait for papaya breeding due to its relationship with production and use as a standard quality of fruits. Fruit mass can be determined from volume, if its density is known. In this study, the estimated average fruit density is 0.85 g that was determined by the equation 2. The fruit mass (FM) estimated trought the IP methodology, using the equation 3 is not different (p >0.05) from the actual fruit mass estimated trought analytical scale (Table 3).

The average FM determined through the IP methodology was
Table 1. Estimates of averages, coefficient of variation (CV), average difference confidence interval (CI), Pearson correlations (R) for the evaluated traits using image-based phenotyping (IBP) and manual phenotyping (MP) methodologies in ‘THB’ cultivar.

<table>
<thead>
<tr>
<th>Trait</th>
<th>Average MP</th>
<th>Average IBP</th>
<th>CV MP</th>
<th>CV IBP</th>
<th>p value</th>
<th>CI</th>
<th>R</th>
<th>Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>FL</td>
<td>13.49</td>
<td>13.47</td>
<td>6.78</td>
<td>6.75</td>
<td>0.232</td>
<td>[-0.0133 ; 0.0535]</td>
<td>0.99</td>
<td>0.15</td>
</tr>
<tr>
<td>FD</td>
<td>8.26</td>
<td>8.23</td>
<td>5.77</td>
<td>5.42</td>
<td>0.144</td>
<td>[-0.0122 ; 0.0797]</td>
<td>0.94</td>
<td>0.36</td>
</tr>
<tr>
<td>OCL</td>
<td>9.63</td>
<td>9.61</td>
<td>8.84</td>
<td>8.76</td>
<td>0.187</td>
<td>[-0.0106 ; 0.0528]</td>
<td>0.99</td>
<td>0.21</td>
</tr>
<tr>
<td>OCD</td>
<td>4.26</td>
<td>4.23</td>
<td>9.66</td>
<td>9.21</td>
<td>0.060</td>
<td>[-0.0008 ; 0.0685]</td>
<td>0.95</td>
<td>0.70</td>
</tr>
<tr>
<td>TP</td>
<td>2.04</td>
<td>2.03</td>
<td>8.68</td>
<td>8.39</td>
<td>0.061</td>
<td>[-0.0006 ; 0.0267]</td>
<td>0.96</td>
<td>0.49</td>
</tr>
</tbody>
</table>

FL: Fruit length expressed in cm; FD: Fruit diameter expressed in cm; OCL: Ovarian cavity length expressed in cm; OCD: Ovarian cavity diameter expressed in cm; TP: Average thickness Pulp expressed in cm.

Fig 1. A) Correlation between estimated fruit volume values by water displacement method (WDM) and the volume estimated by the equation 1 using dimensions performed with image processing (IP); B) Bland-Altman plot for the comparison of papaya volumes measured with water displacement method (WDM) and volume estimated using the dimensions performed with image processing (IP); outer lines indicate 95% limits of agreement (-62.88 mL; 49.24 mL), while center line showing the average difference (-7 mL).

Table 2. Estimates of averages, coefficient of variation (CV), average difference confidence interval (CI), Pearson correlations (R) for the evaluated fruit volume (FV) trait using image-based phenotyping (IBP) and manual phenotyping (MP) methodologies in ‘THB’ cultivar.

<table>
<thead>
<tr>
<th>Trait</th>
<th>Average MP</th>
<th>Average IBP</th>
<th>CV MP</th>
<th>CV IBP</th>
<th>p value</th>
<th>CI</th>
<th>R</th>
<th>Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>FV</td>
<td>473.50</td>
<td>480.32</td>
<td>17.99</td>
<td>15.45</td>
<td>0.098</td>
<td>[-14.948 ; 1.310]</td>
<td>0.95</td>
<td>1.42</td>
</tr>
</tbody>
</table>

Fig 2. Correlation between the estimated papaya fruit mass through the equation 3 using dimensions performed with image
processing (IP), and actual papaya fruit mass measured in analytical scale. **Table 3.** Estimates of averages, coefficient of variation (CV), average difference confidence interval (CI), Pearson correlations (R) for the evaluated fruit mass (FM) trait using image-based phenotyping (IBP) and manual phenotyping (MP) methodologies in ‘THB’ cultivar.

<table>
<thead>
<tr>
<th>Trait</th>
<th>Average MP</th>
<th>Average IBP</th>
<th>CV MP</th>
<th>CV IBP</th>
<th>p value</th>
<th>CI</th>
<th>R</th>
<th>Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>FM</td>
<td>395.16</td>
<td>395.76</td>
<td>17.45</td>
<td>17.12</td>
<td>0.699</td>
<td>[-3.682 ; 2.486]</td>
<td>0.99</td>
<td>0.15</td>
</tr>
</tbody>
</table>

**Fig 3.** Image capturing system. (1) laptop computer with processor Intel(R) Celeron(R) CPU B800 @ 1.50GHz @ 1.50GHz, memory (RAM) 8000 GB, System type: 64-bit operating system, and (2) flat-bed scanner model Optco-Pro A320.

**Table 4.** Measured traits in two groups of papaya fruits used in this study.

<table>
<thead>
<tr>
<th>Traits</th>
<th>First group</th>
<th>Second group</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>MP</td>
<td>IBP</td>
</tr>
<tr>
<td>Fruit length</td>
<td>Digital caliper</td>
<td>IP</td>
</tr>
<tr>
<td>Fruit diameter</td>
<td>Digital caliper</td>
<td>IP</td>
</tr>
<tr>
<td>Ovarian cavity length</td>
<td>Digital caliper</td>
<td>IA</td>
</tr>
<tr>
<td>Ovarian cavity diameter</td>
<td>Digital caliper</td>
<td>IA</td>
</tr>
<tr>
<td>Thickness pulp</td>
<td>Digital caliper</td>
<td>IA</td>
</tr>
<tr>
<td>Density</td>
<td>Formula</td>
<td>-</td>
</tr>
<tr>
<td>Fruit volume</td>
<td>WDM</td>
<td>IP</td>
</tr>
<tr>
<td>Fruit mass</td>
<td>AB</td>
<td>IP</td>
</tr>
</tbody>
</table>

IBP: image-based phenotyping; MP: manual phenotyping; IP: image processing; IA: image analysis; WDM: water displacement method; AB: analytical scale.

**Fig 4.** Flow diagram of image processing for the macro Papaya measurements traits. The macro is recording of image processing by ImageJ software that automates image processing without the need to perform all the steps from importing the image sequence to the measurement of the traits. So, in the software, the macro ‘Papaya measurements imports the image sequences in the RGB format of a specific folder of the computer, calibrates the software using the ‘set scale’ tool, then the macro transforms the images into ‘8-bit’ images and these into binary images using the ‘make binary’ tool. After processing the images, the software made the measurement of the traits (area, length, diameter, perimeter ...) through the tool ‘Analyze particles’. The results are automatically
 Exported to a Microsoft Excel worksheet.

Fig 5. Digital image processing with ImageJ software. (a) Original RGB color image of papaya fruit; (b) eight-bit grayscale image, and (c) two-bit binary image of papaya fruit.

395.16 g and the actual average FM was 395.76 g. The estimate error using digital images was 0.15 % for the FM trait, showing that IP method can be deployed to estimate this trait with high degree of accuracy (Table 4).

The correlation between the actual FM and the fruit mass estimated through the IP methodology is shown in Fig. 2, in which the intersection has pointed the actual FM to the estimated FM. They have evenly distributed along the line, implying high correlation (R = 0.99) and good consistency between methodologies. Image-based phenotyping was an efficient phenotypic analysis instrument when compared to the MP. These methodologies based on digital images are promising tools that can be used to assist the phenotyping fruit traits in the papaya crop. The images are captured with a flat-bed scanner, which is simple, inexpensive and easily handled and transported. In addition, once stored in the computer, the images can be accessed for later analysis. The use of images reduces the time for collecting and analyzing data, increasing the efficiency in phenotyping papaya fruit traits. In this study an observer 3600 s was used for phenotyping fruit traits using the MP, while the same observer that was used for IBP employed 900 s to estimate the same traits. Comparing the two methodologies, phenotyping using IBP is about four-times quicker. The expectation is that IBP methodologies allow for expanding the experiments, performing phenotypic evaluation quickly and precisely, contributing to increases in selection differential and the heritability coefficient with a direct effect on genetic gain.

Materials and methods

Image-based phenotyping (IBP) methodology

Image capturing

The image capturing system consisted of a flat-bed scanner model Optico-Pro A320, with USB connection and a laptop computer with processor Intel(R) Celeron(R) CPU B800 @ 1.50GHz @ 1.50GHz, memory (RAM) 8000 GB, System type 64-bit operating system, equipped with ImageJ Software (Version 1.50c) (Fig. 3). The fruits were sliced into half longitudinally and each fruit (half) was placed on a flat-bed scanner and scanned to create the digital image. Images were saved in the .jpg compressed format at 100 dpi resolution.

Image analysis (IA) methodology

For the phenotyping using image analysis (IA), only RGB images were used. The image sequence of the fruits was imported into ImageJ software and software calibration was performed through the ‘set scale’ tool using an image with a known measuring ruler. Then the measurements of the traits were performed using the ‘straight line selection’ tool (Table 4).

Image processing (IP) methodology

Papaya measurements traits macro was created to the IP automate Fig. 4. This macro opens the image sequence and segments of images and automatically performs measurements of the traits. When opening ImageJ software and select the ‘Papaya measurements traits’ the macro opens the image sequences in a specific folder located in computer in original RGB color (Fig. 5a). After reading the image sequence in RGB format, each image was converted to an eight-bit grayscale image (Fig. 5b). Using the thresholding technique, the region of interest in the grayscale image was segmented using the Otsu algorithm. After that, a binary image with pixels’ values of 0 (black) or 250 (white) is obtained (Fig. 5c). From the binary images and using ‘analyze particles’ tool, it was possible to perform the measurements. Using this tool, the software releases a spreadsheet with the measurements data.

Plant material

Papaya fruits of ‘THB’ cultivar were randomly sampled in different points from the 34-commercial area, at Caliman Agrícola S.A., in Linhares, Espírito Santo, Brazil (19°06’ and 57°103’) and 39°45’ and 40°19’W). Two fruits per plant were collected in 50 plants of the 34 area, totaling 100 fruits. For the evaluation the papaya fruits were divided into two groups with 50 fruits each, and all fruits presented zero maturation stage. The first group was used to validate the image-based phenotyping (IBP) methodology. For the validation, the fruit traits were measured using the manual methodology (MM) traditionally used in the crop and the IBP methodology (image analysis - IA and image processing - IP). In the second group the FV and FM were measured using only the image processing (IP).
**Fruit traits phenotyping in the first group**

The thickness pulp (TP), ovarian cavity length (OCL) and ovarian cavity diameter (OCD) were measured using IA. For the traits fruit length (FL), fruit diameter (FD), fruit volume (FV) and fruit mass (FM) were measured using IP. These traits also were manually phenotyped using a digital caliper model Digimess IPS4 and an analytical scale model Toledo 9094, respectively (Table 4). The traits FL and FD measured using the IP. They were used to estimate the FV according to the equation adapted from (Koc, 2007):

\[ V = \frac{\pi \times FL \times FD^2}{6} \]  
(1)

Where; \( V \) is the estimated fruit volume (mL); \( FL \) is the fruit length (cm) and \( FD \) is the fruit diameter (cm).

To validate the estimated fruit volume through the IP, the volume of each fruit was calculated through water displaced method (WDM). In this method, each fruit was immersed in a graduated volumetric container of 10 L containing an initial known volume of water (initial volume); the fruit volume was obtained by calculating the difference between the final and initial volume.

Using of FV obtained by the WDM and FM measured by analytical scale, we estimated the average density of the fruits through the following equation:

\[ ad = \frac{FM}{Ve} \]  
(2)

Where; \( ad \) is the average density (g mL\(^{-1}\)), \( Ve \) is the fruit volume (mL), and \( FM \) is the fruit mass (g).

**Fruit traits phenotyping in the second group**

In this group it was estimated the FL, FD and FV traits using the IP. With the volume estimated through IP it was possible to estimate the fruit mass using the following equation:

\[ FM_e = Ve \times ad \]  
(3)

Where; \( FM_e \) is the estimated fruit mass in the second group (g), \( Ve \) is the estimated fruit volume in the second group (mL) and \( ad \) is the average fruit density (g mL\(^{-1}\)) obtained in the first group. In order to validate the FM estimate using the IP, the fruits were weighed in analytical scale.

**Statistical analysis**

The paired t-test and the average difference confidence interval approach were used to compare the FL, FD, OCL, OCD, TP and FM traits, using the IP and MP methodologies. The paired t-test was also used to compare the volume determined from the WDM and IP.

Also, the Bland-Altman approach was used to plot the agreement between the calculated and measured papaya volume (Bland and Altman 1999). In addition, the coefficients of variation and confidence intervals were estimated for these traits. The interest was to determine whether the image processing methodology differs statistically from manual methodology traditionally used. The estimated FM using the IP was compared to FM obtained by analytical scale through the Pearson correlations. The statistical analyses were performed using SAS Studio (SAS Institute, Cary, NC, USA) program.

The average relative error of the digital images methodologies was calculated according to the following equation (Zhang, 2000):

\[ Error(\%) = \left| \frac{x_m - x_i}{x_m} \right| \times 100 \]  
(4)

Where: \( x_m \) is the average of the trait obtained using the manual methodology, and \( x_i \) is the average of the trait obtained by digital image.

**Conclusion**

The image-based phenotyping proposed in this study was more efficient compared to manual phenotyping, man labor and precision terms. The IBP proposed in this study, allowed the quick and precise phenotyping of morphological traits in papaya fruits. It can be incorporated in papaya breeding program routine, contributing to the increase in the differential selection with a direct effect on the genetic gains.

The IBP method used in this study can be applied in other fruits tree.

**Acknowledgments**

This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES) –Finance Code 001. The authors kindly thank the North Fluminense State University “Darcy Ribeiro” (UENF) and Caliman Agricola S.A. Company, for the support.

**References**


