



Image analysis of seeds and machine learning as a tool for distinguishing populations: Applied to an invasive tree species

Francival Cardoso Felix^{1*}, Kyvia Pontes Teixeira das Chagas¹, Fernando dos Santos Araújo², Josenilda Aprigio Dantas de Medeiros³, Fábio de Almeida Vieira³, Salvador Barros Torres⁴ and Mauro Vasconcelos Pacheco³

¹Universidade Federal do Paraná, Rua XV de Novembro, 1299, 80060-000, Curitiba, Paraná, Brazil. ²Universidade Federal do Agreste de Pernambuco, Garanhuns, Pernambuco, Brazil. ³Universidade Federal do Rio Grande do Norte, Macaíba, Rio Grande do Norte, Brazil. ⁴Universidade Federal Rural do Semi-Árido, Mossoró, Rio Grande do Norte, Brazil. *Author for correspondence. E-mail: felixfc@outlook.com.br

ABSTRACT. Invasive species threaten crops and ecosystems worldwide. Therefore, we sought to understand the relationship between the geographic distribution of species populations and the characteristics of seeds using new techniques such as seed image analysis, multivariate analysis, and machine learning. This study aimed to characterize *Leucaena leucocephala* (Lam.) de Wit. seeds from spatially dispersed populations using digital images and analyzed their implications for genetic studies. Seed size and shape descriptors were obtained using image analysis of the five populations. Several analyses were performed including descriptive statistics, principal components, Euclidean distance, Mantel correlation test, and supervised machine learning. This image analysis technique proved to be efficient in detecting biometric differences in *L. leucocephala* seeds from spatially dispersed populations. This method revealed that spatially dispersed *L. leucocephala* populations had different biometric seed patterns that can be used in studies of population genetic divergence. We observed that it is possible to identify the origin of the seeds from the biometric characters with 80.4% accuracy (Kappa statistic 0.755) when we applied the decision tree algorithm. Digital imaging analysis associated with machine learning is promising for discriminating forest tree populations, supporting management activities, and studying population genetic divergence. This technique contributes to the understanding of genotype-environment interactions and consequently identifies the ability of an invasive species to spread in a new area, making it possible to track and monitor the flow of seeds between populations and other sites.

Keywords: artificial intelligence; descriptors; Euclidean distance; Mantel correlation; seed analysis; *Leucaena leucocephala*.

Received on February 23, 2022.

Accepted on July 18, 2022.

Introduction

Leucaena leucocephala (Lam.) de Wit. (Fabaceae) is a tree native to Central America and Mexico. It is currently distributed in most tropical areas of the world (Fonseca & Jacobi, 2011) and occurs in cultivated and invaded areas (Crawford et al., 2015). The occurrence of *L. leucocephala* is limited to temperatures between 10 and 40°C with annual rainfall between 600 and 1,700 mm. However, this species can also occur in regions with only 250 mm of annual rain and drought for up to eight months, as is the case in semiarid conditions (Drumond & Ribaski, 2010).

Its growth is relatively fast, making it promising for exploration in semiarid regions (Walker, 2012), which justifies the work done with this species in dry forests. In addition, its potential is associated with its ability to regrow, resist drought, adapt to soil conditions, productivity, and acceptance by animals as food (Cámara et al., 2015; Crawford et al., 2015; Azuara-Morales et al., 2020; Barros et al., 2020; Dueñas et al., 2020). The energy potential of its wood is similar to that of *Eucalyptus grandis* W. Hill (ex Maiden) (Machado, Andrade, Silva, Sena, & Thode-Filho, 2014), whose calorific capacity is adequate for charcoal production (Silva et al., 2018).

Invasive *L. leucocephala* has high potential for expansion in biodiversity hotspots and can advance in vulnerable and critical areas, including natural reserves in tropical and subtropical forests (Wan & Wang, 2018). There was an abundance of *L. leucocephala* semi-domesticated populations that spread spontaneously in marginal and disturbed areas. These may have different adaptive characteristics, depending on their

geographical origins. Exotic species initially encounter adaptive difficulties when colonizing new environments, resulting from less genetic variation. However, the introduction of different genetic lineages drives the natural selection of genotypes that are more adapted over time, thus generating effective colonization and consequently increasing their potential for invasion (Wang et al., 2012).

For instance, *Ambrosia artemisiifolia* L., native to North America, was introduced in the Rhône Valley, France during the 18th century. Currently, it is considered to be a highly invasive species in this region. According to a study, the introduction occurred from several sources, which explains the high genetic variability of *A. artemisiifolia* populations in France, reaching even more significant intrapopulation genetic variability than in populations of North American origin (Genton, Shykoff, & Giraud, 2005).

In this sense, the characterization of populations from different geographical origins is relevant to understanding the adaptive potential of plant species and to apply the results to genetic improvement programs. Therefore, the biometric study patterns of seeds from different populations can be an efficient method for this purpose, as they are the primary vehicle for the spread of germplasm. Variations in seed biometrics are intricately linked to environmental alterations and the genetic diversity of populations, which may be associated with their geographical origin (Brus et al., 2011; Santana, Torres, & Benedito, 2013; Costa et al., 2016; Khumaeva, Khabibov, & Muratchaeva, 2016; Rewicz, Bomanowska, Magda, & Rewicz, 2016; Pontes et al., 2018; Alfaro-Solís et al., 2020; Felix et al., Medeiros, Ferrari, Vieira, & Pacheco, 2020).

Traditionally, seed biometry is performed to measure seed dimensions manually using a caliper (Oliveira, Araújo-Neto, Cruz, & Ferreira, 2012; Santana et al., 2013; Costa et al., 2016; Dutra et al., 2017; Menegatti, Mantovani, Navroski, & Souza, 2017; Roveri-Neto & Paula, 2017; Silva, Cardoso, Dutra, & Morais, 2017; Santos et al., 2018; Correia, Felix, Araujo, Ferrari, & Pacheco, 2019; Zuffo et al., 2019). However, this method requires a lot of time to obtain a sufficient sample size and is susceptible to human errors. These drawbacks can be solved using the image analysis technique, as it allows for the quantification of several descriptors associated with the seed surface (size, shape, and color) in a faster and more efficient way than traditional methods (Dell'Aquila, 2006; Varma, Durga, & Keshavulu, 2013). Additionally, seed image analysis can distinguish species within the same genus (Bacchetta, Grillo, Mattana, & Venora, 2008; Bianco et al., 2016), providing information on identifying genotypes from different locations (Medina, Skurtys, & Aguilera, 2010; Rewicz et al., 2016), assisting in taxonomy and species delimitation (Bianco et al., 2016; Uccesu et al., 2016), seed morphological differentiation (Orrù, Grillo, Lovi, Venora, & Bacchetta, 2012), and seeds morphometry (Noronha, Pereira, Flores, Demartelaere, & Medeiros, 2018).

Artificial intelligence and machine learning can be widely used to recognize seeds based on the characteristics of each species (Bao & Bambil, 2021) and genotype (Tan, Wang, Li, & Gong, 2019; Landa et al., 2021; Ropelewska & Piecko, 2022). We hypothesize that there is congruency between the biometric descriptors of seeds among populations and their geographical distances. Additionally, we sought to understand how diversity of the seed descriptors of *L. leucocephala* were obtained for its invasive range. The paper aimed to characterize *L. leucocephala* seeds from spatially dispersed populations through digital imaging and analyze their implications for genetic studies.

Material and methods

Characterization of populations

We selected five newly established *L. leucocephala* populations in Rio Grande do Norte State, Brazil (Figure 1). Each feral population (escaped cultivation) was delimited on local roads. The sampled populations were fitted to a group of individuals corresponding to an area greater than 1.0 ha. The fruits (pods) were manually harvested from five individuals per population, and seed processing consisted of manual removal of pods and selection of healthy seeds, disregarding those that were defective and damaged by insects.

Image processing

The seeds of each population were subjected to digital image processing to obtain biometric descriptors. Seed biometrics were analyzed using the digital image editing program ImageJ® (<https://imagej.nih.gov/ij/>). For each population, 800 seeds were photographed (12 MP camera) from 20 cm away on a matte white paper background, with a ruler graduated in millimeters. Image processing was based on converting the original image (Figure 2A) into an 8-bit format (256 tones) (Figure 2B), followed by calibration of the scale in

millimeters, selection of the area to be analyzed, and inversion of the gray spectrum (Figure 2C). We then applied the 'threshold' mask to differentiate between the image components (Figure 2D) and particle analysis, thus obtaining biometric descriptors (Figure 2E), whose data were exported to an electronic data sheet (Microsoft Office Excel®).



Figure 1. *Leucaena leucocephala* populations from the Eastern Rio Grande do Norte State, Brazil [Populations: 1 - São José de Mipibú County (6°00'23.84" S and 35°15'59.84" W), 2 - Macaíba County (5°53'02.63" S and 35°21'49.13" W), 3 - São José de Mipibú County (6°03'30.62" S and 35°18'29.09" W), 4 - São José de Mipibú County (6°00'6.35" S and 35°15'50.20" W), and 5 - Natal County (5°50'32.84" S and 35°12'2.37" W)].

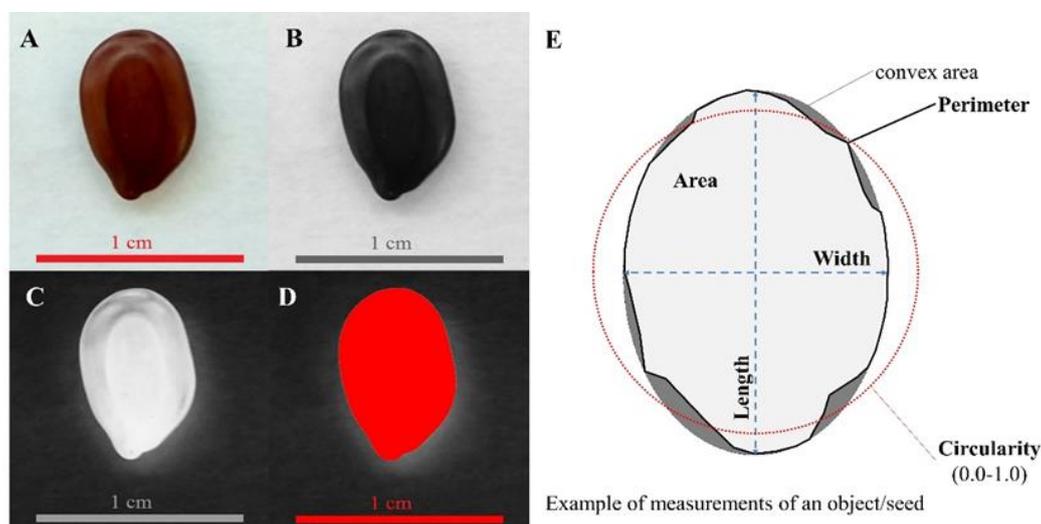


Figure 2. Steps in digital image processing to obtain biometric descriptors in *Leucaena leucocephala* seeds. A: photo capture; B: conversion to 8-bit format; C: inversion of the gray spectrum; D: use of the 'threshold' mask; E: measurements of the seed perimeter, area, solidity (area/convex area); length, width, aspect ratio (major axis/minor axis), circularity ($4\pi \frac{[\text{area}]}{[\text{perimeter}]^2}$) and roundness

($4 \frac{[\text{area}]}{\pi \cdot [\text{major axis}]^2}$) were calculated automatically.

The evaluated biometric descriptors followed the descriptions of Ferreira and Rasband (2012):

Area - seed selection area in calibrated square units, calculated within the perimeter defined polygon (mm²);
Perimeter - length of the external limits of seed selection in calibrated units, calculated from the center of the limit pixels (mm);

Circularity - scalar value from 0.0 to 1.0, indicating a perfect circle when close to 1.0, for the seed shape in relation to its perimeter $4\pi \frac{[\text{area}]}{[\text{perimeter}]^2}$, and an elongated shape when close to zero;

Length - greater distance between two points along the seed selection limit in calibrated units (mm);

Width the smallest distance between two points along the seed selection limit in calibrated units (mm);

Aspect ratio - proportion of the adjusted seed ellipse from the major and minor axes $\frac{[\text{major axis}]}{[\text{minor axis}]}$;

Roundness - scalar value from 0.0 to 1.0, inverse of the aspect ratio $4 \frac{[\text{area}]}{\pi \cdot [\text{major axis}]^2}$;

Solidity - scalar value from 0.0 to 1.0, indicated concerning seed convex area selection $\frac{[\text{area}]}{[\text{convex area}]}$. The convex area is characterized as a delimiter of the original shape of the object, such as a rubber band tightly wrapped around the area (Matsumoto et al., 2015).

Data analysis

The experimental design was completely random, with 800 sample units for each *L. leucocephala* population. We proceeded to the analysis of descriptive statistics, and cluster analysis by Euclidean distance was used to verify similarities between populations with the elaboration of a hierarchical dendrogram of the unweighted pair group method with arithmetic mean (UPGMA) similarity with the standardized group mean for the biometric descriptors. The relationship between spatial distance and population dissimilarity was analyzed using the Mantel correlation test (r_M), with 10,000 simulations between Euclidean distance and geographic distance ($p < 0.05$). The statistical program used was BioEstat® (version 5.3) (<https://www.mamiraua.org.br/downloads/programas/>).

Multivariate analysis of the principal components (PCA) for the biometric descriptors was performed with standardized data ($\frac{x - \text{mean}}{\text{standard deviation}}$) to obtain eigenvalues and variance for the contribution of each principal component, eliminating the components with few representations. The statistical program used was Past® (version 3.20) (https://palaeo-electronica.org/2001_1/past/issue1_01.htm). In addition, a machine-learning algorithm supervised by decision trees (J48-C0.25-M2) was tested. The program used was Weka® (version 3.8) (<https://www.cs.waikato.ac.nz/ml/weka/>).

Results

The seeds of *L. leucocephala* had different physical characteristics between populations, as evidenced by the biometric descriptors quantified in the digital processing of the images for seed shape and size (Table 1). Seeds from population POP1 were larger in area ($33.85 \pm 4.03 \text{ mm}^2$), perimeter ($22.65 \pm 1.47 \text{ mm}$), length ($8.34 \pm 0.59 \text{ mm}$) and width ($5.42 \pm 0.34 \text{ mm}$). In comparison, these characteristics were smaller in seeds from POP2 (area: $22.07 \pm 2.27 \text{ mm}^2$, perimeter: $18.17 \pm 1.03 \text{ mm}$, length: $6.70 \pm 0.44 \text{ mm}$, and width: $4.47 \pm 0.24 \text{ mm}$). The characteristics of the seeds of the other populations occupied an intermediate position between POP1 and POP2, proving physical differences between populations in the seeds of this species.

Table 1. Biometrics of seeds from different populations of *Leucaena leucocephala* analyzed using a digital image (values are written as the mean \pm standard deviation).

Descriptors	Populations					Average
	POP1	POP2	POP3	POP4	POP5	
Area (mm ²)	33.85±4.03	22.07±2.27	29.24±2.86	30.35±5.36	22.79±4.83	27.66
Perimeter (mm)	22.65±1.47	18.17±1.03	21.42±1.11	21.55±2.06	18.29±2.08	20.42
Circularity (0.00-1.00)	0.827±0.02	0.838±0.02	0.799±0.02	0.815±0.03	0.847±0.02	0.825
Length (mm)	8.34±0.59	6.70±0.44	8.17±0.49	8.12±0.84	6.71±0.84	7.61
Width (mm)	5.42±0.34	4.47±0.24	4.74±0.29	4.92±0.46	4.56±0.47	4.82
Aspect ratio	1.520±0.10	1.478±0.09	1.715±0.13	1.632±0.15	1.450±0.13	1.559
Roundness (0.00-1.00)	0.661±0.04	0.679±0.04	0.587±0.05	0.618±0.06	0.695±0.06	0.648
Solidity (0.00-1.00)	0.972±0.01	0.968±0.01	0.971±0.01	0.972±0.01	0.970±0.01	0.971

Populations: POP1 - São José de Mipibú, POP2 - Macaíba, POP3 - São José de Mipibú, POP4 - São José de Mipibú, and POP5 - Natal County.

When examining the characteristics of the *L. leucocephala* seed proportion and the ratio of the largest and smallest plane of the particle evaluated in the image, it can be seen that the seeds of POP3 were the largest (1.715 ± 0.13) and those of POP5 were the smallest (1.450 ± 0.13) (Table 1). The seeds of POP5 showed greater circularity (0.847 ± 0.02) and roundness (0.695 ± 0.06), whereas those of POP3 were less circular (0.799 ± 0.02) and round (0.587 ± 0.05), precisely because of the higher proportion of length and width. The solidity of the seeds of POP1 and POP4 had the highest index (0.972 ± 0.01), whereas those of POP2 had the lowest index (0.968 ± 0.01). The high solidity values indicate that the seed shape has low irregularity on its outer edges.

The first two principal components were selected because they represented 90.6% of the data variation [Components 1 (60.9%) and 2 (29.7%)] (Figure 3). The seed width, length, perimeter, area, and solidity were highly correlated in the second quadrant (Q2), whereas shape variables of roundness and circularity were highly correlated in the first quadrant (Q1), except for aspect ratio (Q3). Therefore, measurements of the area, perimeter, length, or width were fundamental in capturing most of the existing variations in seed size, whereas circularity and aspect ratio or roundness were important for the seed shape. The seeds of the different populations were widely spread, but it was possible to verify the concentration of each population (different colors).

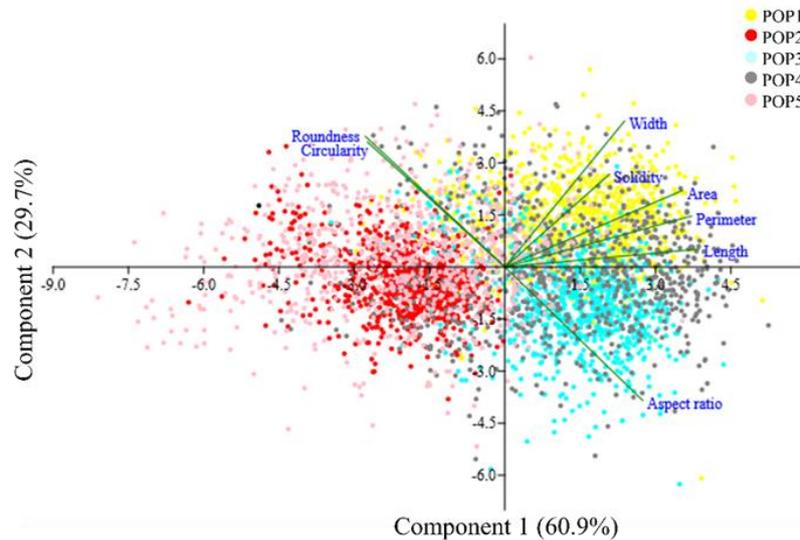


Figure 3. Principal component analysis (PC1 x PC2) based on the digital biometric analysis of *Leucaena leucocephala* seeds.

The clustering of populations according to the Euclidean distance distinguished the formation of two groups (group I: POP1, POP3, and POP4; group II: POP2 and POP5) at a cut-off of 70% (Figure 4). Euclidean distance analysis showed that populations POP2 and POP5 were more similar to each other (1.552), while POP1 and POP2 were more divergent (5.788) according to the biometric descriptors evaluated (Table 2). A positive and significant Mantel correlation ($r_M = 0.613$, $p < 0.05$) was observed between the biometric descriptors of *L. leucocephala* seeds and the geographic distance of the studied populations. Additionally, we observed that it was possible to identify the seed origin of *L. leucocephala* populations from the analysis of the biometric traits with 80.4% accuracy (Kappa statistic 0.755) when we applied the decision tree algorithm. The accuracy of correct seed identification ranged from 75.4 to 84.3% (Table 3).

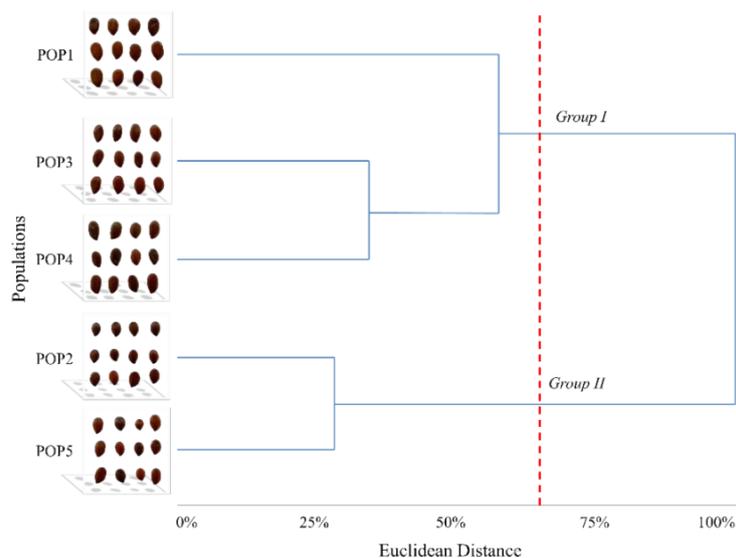


Figure 4. Cluster analysis based on Euclidean distance for five populations of *Leucaena leucocephala* and their biometric descriptors (formation of groups at 70% cut-off level).

Table 2. Euclidean distance and geographic distance between five *Leucaena leucocephala* populations.

Euclidean distance	Geographic distance				
	POP1	POP2	POP3	POP4	POP5
POP1	---	17,290 m	7,350 m	610 m	19,570 m
POP2	5.788	---	20,250 m	17,070 m	18,630 m
POP3	4.008	5.524	---	7,950 m	26,690 m
POP4	2.485	5.089	1.730	---	18,960 m
POP5	5.229	1.552	5.655	4.808	---

Populations: POP1 - São José de Mipibú, POP2 - Macaiba, POP3 - São José de Mipibú, POP4 - São José de Mipibú, and POP5 - Natal County. m: distance in meters.

Table 3. Confusion matrix and precision correct identification of seeds from five *Leucaena leucocephala* populations using machine learning.

Classification	POP1	POP2	POP3	POP4	POP5	Precision correct identification
POP1	738	1	23	23	15	80.7%
POP2	5	714	4	10	67	75.4%
POP3	26	12	723	32	7	80.8%
POP4	116	45	123	493	23	84.3%
POP5	29	175	22	27	547	83.0%

Populations: POP1 - São José de Mipibú, POP2 - Macaiba, POP3 - São José de Mipibú, POP4 - São José de Mipibú, and POP5 - Natal County.

Discussion

This study evaluated the seed traits of spatially dispersed populations using digital images of *L. leucocephala* and analyzed their implications for genetic studies. We found that distant populations differed in their seed traits. Diversity in the shape and size of seeds is related to the genetic characteristics of the populations and habitats in which they were found. The phenotypic diversity of *L. leucocephala* seeds can be attributed to the pantropical distribution of the species (Fonseca & Jacobi, 2011; Crawford et al., 2015; Wan & Wang, 2018) which favors rapid establishment in new environments.

Related studies have demonstrated an association between the biometric characteristics of fruits and seeds and genotype distinctions, as demonstrated in this study. For instance, different populations of *Myrtus communis* L. in Palermo, Italy were evaluated for genetic diversity and morphological characteristics of fruits and seeds. Two of the main groups were significantly correlated with biometric characteristics, which can be used as morphological markers for fruit production in selection and breeding programs (Melito et al., 2016). In a study of the morphological diversity of *Argania spinosa* Skeels, it was possible to differentiate genotypes based on the characteristics of the fruits for preservation *in situ* in the Admine Reserve, Morocco (Metougui, Mokhtari, Maughan, Jellen, & Benlhabib, 2017).

The establishment of exotic species in new areas increases population distribution. It contributes to the genetic divergence among individuals (Wan & Wang, 2018), as reported in the Rhône Valley, France. In this study, populations of *A. artemisiifolia* showed high intrapopulation genetic diversity in the original introduction area, with a reduction in the range of natural expansion/colonization of the species from sequential bottlenecks; which was expected at this stage of expansion when superior genotypes were selected (Genton et al., 2005).

Biometrics of forest seeds have been reported with sample numbers equal to or greater than 100 seeds (Santana et al., 2013; Costa et al., 2016; Dutra et al., 2017; Menegatti et al., 2017; Roveri-Neto & Paula, 2017; Silva et al., 2017; Noronha et al., 2018; Correia et al., 2019; Zuffo et al., 2019; Rosa et al., 2020). This is because of the difficulty in analyzing seeds using traditional methods; therefore, the image processing technique is a suitable alternative for the analysis of many samples. In addition, the biometric analysis of seeds using images also provides further information related to the size and shape of seeds, whereas the manual method using digital calipers is limited to analyzing only length, width, and thickness (Felix et al., 2020). Thus, digital image processing allows for the exploration of several seed descriptors if there is a reference plane for the captured images.

Biometric analysis of seeds can be applied to studies on genetic divergence and variability between populations and individuals (Cruz, Regazzi, & Carneiro, 2012; Menegatti et al., 2017; Roveri-Neto & Paula, 2017; Pontes et al., 2018; Correia et al., 2019; Alfaro-Solís et al., 2020; Felix et al., 2020; Rosa et al., 2020). Furthermore, they are considered fundamental for population dynamics studies because they provide information on the establishment of species in new environments (Brus et al., 2011; Costa et al., 2016; Khumaeva et al., 2016; Rewicz et al., 2016; Dutra et al., 2017; Silva et al., 2017; Santos et al., 2018). The variations found in biometric studies are closely related to genetic factors (Santana et al., 2013), since during

the maturation seed process, the dimensions of the seed are determined genetically (Bewley, Bradford, Hilhorst, & Nonogaki, 2013).

The results obtained through seed imaging and multivariate analysis showed that seeds provide important information about the genetic divergence between *L. leucocephala* populations. This is because multivariate analyses were used, which, through simplified functions, demonstrate trends and variations in the data to estimate the divergence between genotypes (Costa et al., 2016). The two principal components were adequate because they represent 90.6% of the data variation, taking into account the assumptions of having an eigenvalue greater than 1.0 (Kaiser, 1960), and concentrating at least 70% of the total variation between variables (Medina et al., 2010).

Clustering methods were efficient in verifying the existence of genetic variability and similarity between genotypes as well as the principal component to concentrate the number of variables to facilitate the interpretation of results. The results obtained through seed image analysis and multivariate analysis showed that seeds provide important information about the adaptation and evolution of *Anchusa* L. taxa in Sardinia (Farris et al., 2020), and helped to identify seeds from different locations (Pan et al., 2021; Cecco, Musciano, D'Archivio, Frattaroli, & Martino, 2019), and the distribution of species from Central Europe (Mazur, Marcysiak, Dunajska, Gawlak, & Kałuski, 2022). In the present study, seeds from closer populations also presented similar biometric characteristics.

A significant Mantel correlation suggests an isolation-by-distance pattern. In other words, the higher the geographical distance, the more divergent the seeds were with respect to the evaluated parameters. Individuals of *L. leucocephala* do not reach a large size, are considered a pioneer heliophyte species of autochoric and zoochoric reproduction, and can grow in stress conditions that limit its expansion in non-degraded or disturbed areas (Costa & Durigan, 2010). Therefore, there were divergences between the assessed populations based on the biometric descriptors of the seeds, and geographic distribution influenced the variability between populations of *L. leucocephala*. Our results suggest that the analysis of seed images can also be tested in other forest species, aiming to estimate the degree of genetic variability between native, exotic, or invasive populations.

The results of machine learning demonstrated the effectiveness of recognizing seeds from different populations using biometric size and shape descriptors. Finally, the use of machine learning automates the recognition of seeds from distinct geographic origins, which has also been proven for *L. leucocephala*. In view of this, new studies at the continental level or between biogeoclimatic zones will contribute to the understanding of the expansion of invasive species that cause agricultural production losses and invade forest ecosystems.

Conclusion

Image analysis was efficient in detecting biometric differences in *L. leucocephala* seeds from distinct locations. Therefore, this method is promising for discriminating forest tree populations associated with machine learning, supporting management activities, and studying population genetic divergence. Additionally, digital imaging analysis contributes to understanding genotype–environment interactions and, consequently, to identifying the ability of an invasive species to spread in a new area, making it possible to track and monitor the flow of seeds between populations and other sites. In the future, the analysis of seed images can also be tested in other forest species to estimate genetic variability between native, exotic, or invasive populations.

Acknowledgements

We thank the National Council for Scientific and Technological Development (CNPq) for providing a scholarship and financial support to the first and second authors.

References

- Alfaro-Solís, J.D., Montoya-Arroyo, A., Jiménez, V.M., Arnaez-Serrano, E., Pérez, J., Vetter, W., Frank, J., & Lewandowski, I. (2020). *Acrocomia aculeata* fruits from three regions in Costa Rica: An assessment of biometric parameters, oil content and oil fatty acid composition to evaluate industrial potential. *Agroforestry Systems*, 94(1), 1913–1927. DOI: <https://doi.org/10.1007/s10457-020-00511-8>

- Azuara-Morales, I., López-Ortiz, S., Jarillo-Rodríguez, J., Pérez-Hernández, P., Ortega-Jiménez, E., & Castillo-Gallegos, E. (2020). Forage availability in a silvopastoral system having different densities of *Leucaena leucocephala* under Voisin grazing management. *Agroforestry Systems*, *94*(1), 1701-1711. DOI: <https://doi.org/10.1007/s10457-020-00487-5>
- Bacchetta, G., Grillo, O., Mattana, E., & Venora, G. (2008). Morpho-colorimetric characterization by image analysis to identify diaspores of wild plant species. *Flora - Morphology, Distribution, Functional Ecology of Plants*, *203*(8), 669-682. DOI: <https://doi.org/10.1016/j.flora.2007.11.004>
- Bao, F., & Bambil, D. (2021). Applicability of computer vision in seed identification: deep learning, random forest, and support vector machine classification algorithms. *Acta Botanica Brasilica*, *35*(1), 17-21. DOI: <https://doi.org/10.1590/0102-33062020abb0361>
- Barros, V., Melo, A., Santos, M., Nogueira, L., Frosi, G., & Santos, M.G. (2020). Different resource-use strategies of invasive and native woody species from a seasonally dry tropical forest under drought stress and recovery. *Plant Physiology and Biochemistry*, *147*(1), 181-190. DOI: <https://doi.org/10.1016/j.plaphy.2019.12.018>
- Bewley, J. D., Bradford, K. J., Hilhorst, H. W. M., & Nonogaki, H. (2013). *Seeds: physiology of development, germination and dormancy*. New York, NY: Springer.
- Bianco, M. L., Grillo, O., Garcia, P. E., Mascia, F., Venora, G., & Bacchetta, G. (2016). Morpho-colorimetric characterisation of *Malva* alliance taxa by seed image analysis. *Plant Biology*, *19*(1), 1-9. DOI: <https://doi.org/10.1111/plb.12481>
- Brus, R., Ballian, D., Zhelev, P., Pandza, M., Bobinac, M., Acevski, J., ... Jarni, K. (2011). Absence of geographical structure of morphological variation in *Juniperus oxycedrus* L. subsp. *oxycedrus* in the Balkan Peninsula. *European Journal of Forest Research*, *130*(4), 657-670. DOI: <https://doi.org/10.1007/s10342-010-0457-1>
- Câmara, C. S., Alves, A. A., Moreira-Filho, M. A., Garcez, B. S., & Azevêdo, D. M. M. R. (2015). Diets containing leucena or estilosantes hay in lactating mixed-breed Anglo-Nubian goats. *Revista Ciência Agronômica*, *46*(2), 443-450. DOI: <https://doi.org/10.5935/1806-6690.20150025>
- Cecco, V. D., Musciano, M. D., D'Archivio, A. A., Frattaroli, A. R., & Martino, L. D. (2019). Analysis of intraspecific seed diversity in *Astragalus aquilanus* (Fabaceae), an endemic species of Central Apennine. *Plant Biology*, *21*(3), 507-514. DOI: <https://doi.org/10.1111/plb.12844>
- Correia, L. A. S., Felix, F. C., Araujo, F. D. S., Ferrari, C. D. S., & Pacheco, M. V. (2019). Morphometric descriptors and physiological seed quality for selecting *Aspidosperma pyrifolium* Mart. matrix trees. *Revista Caatinga*, *32*(3), 751-759. DOI: <https://doi.org/10.1590/1983-21252019v32n319rc>
- Costa, J. N. M. N., & Durigan, G. (2010) *Leucaena leucocephala* (Lam.) de Wit (Fabaceae): invasive or ruderal? *Revista Árvore*, *34*(5), 825-833. DOI: <https://doi.org/10.1590/S0100-67622010000500008>
- Costa, M. F., Lopes, A. C. A., Gomes, R. L. F., Araújo, A. S. F., Zucchi, M. I., Pinheiro, J. B., & Valente, S. E. S. (2016). Characterization and genetic divergence of *Casearia grandiflora* populations in the Cerrado of Piauí State, Brazil. *Floresta e Ambiente*, *23*(3), 387-396. DOI: <http://dx.doi.org/10.1590/2179-8087.007115>
- Crawford, G., Puschner, B., Affolter, V., Stalis, I., Davidson, A., Baker, T., ... Ostapak, S. (2015). Systemic effects of *Leucaena leucocephala* ingestion on ring tailed lemurs (*Lemur catta*) at Berenty Reserve, Madagascar. *American Journal of Primatology*, *77*(6), 633-641. DOI: <http://dx.doi.org/10.1002/ajp.22386>
- Cruz, C. D., Regazzi, A. J., & Carneiro, P. C. S. (2012). *Modelos biométricos aplicados ao melhoramento genético*. Viçosa, MG: Editora UFV.
- Dell'Aquila, A. (2006). Computerised seed imaging: a new tool to evaluate germination quality. *Communications in Biometry and Crop Science*, *1*(1), 20-31.
- Drumond, M. A., & Ribaski, J. (2010). *Leucena (Leucaena leucocephala): leguminosa de uso múltiplo para o semiárido brasileiro*. Colombo, PR: Embrapa Florestas (Comunicado técnico, 262); Petrolina, PE: Embrapa Semiárido (Comunicado técnico, 142).
- Dueñas, Y. D., Domínguez, J. C. M., Reyes, K. C., Miranda, M. A. R., Quintana, Y. L., & Guanche, L. H. (2020). Potential use of *Leucaena leucocephala* Lam (leucaena) present in agroforestry systems of Pinar del Río. *Revista Cubana de Ciencias Forestales*, *8*(1), 154-162.
- Dutra, F. V., Cardoso, A. D., Bandeira, A. S., Silva, R. M., Morais, O. M., & Prates, C. J. N. (2017). Biometrical characteristics of fruits and seeds of flamboyant. *Scientia Agraria Paranaensis*, *16*(1), 127-132.

- Farris, E., Orrù, M., Uccesu, M., Amadori, A., Porceddu, M., & Bacchetta, G. (2020) Morpho-colorimetric characterization of the Sardinian endemic taxa of the genus *Anchusa* L. by seed image analysis. *Plants*, 9(10), 1-13. DOI: <https://doi.org/10.3390/plants9101321>
- Felix, F. C., Medeiros, J. A. D., Ferrari, C. S., Vieira, F. A., & Pacheco, M. V. (2020). Biometry of *Pityrocarpa moniliformis* seeds using digital imaging: implications for studies of genetic divergence. *Revista Brasileira de Ciências Agrárias*, 15(1), 1-8. DOI: <https://doi.org/10.5039/agraria.v15i1a6128>
- Ferreira, T., & Rasband, W. (2012). *ImageJ: user guide* (ImageJ 1.46r). Bethesda, MA: National Institutes of Health.
- Fonseca, N. G., & Jacobi, C. M. (2011). Germination performance of the invader *Leucaena leucocephala* (Lam.) de Wit. compared to *Caesalpinia ferrea* Mart. ex Tul. and *C. pulcherrima* (L.) Sw. (Fabaceae). *Acta Botanica Brasilica*, 25(1), 191-197. DOI: <https://doi.org/10.1590/S0102-33062011000100022>
- Genton, B. J., Shykoff, J. A., & Giraud, T. (2005). High genetic diversity in French invasive populations of common ragweed, *Ambrosia artemisiifolia*, as a result of multiple sources of introduction. *Molecular Ecology*, 14(14), 4275-4285. DOI: <https://doi.org/10.1111/j.1365-294X.2005.02750.x>
- Kaiser, H. F. (1960). The application of electronic computers to factor analysis. *Educational and Psychological Measurement*, 20(1), 141-151. DOI: <https://doi.org/10.1177/001316446002000116>
- Khumaeva, U. H., Khabibov, A. D., & Muratchaeva, P. M. (2016). Assessment of the structure of the variability of morphological signs of *Trifolium medium* L. introduced under the conditions of mountainous dagestan. *Arid Ecosystems*, 6(1), 135-141. DOI: <https://doi.org/10.1134/S2079096116020074>
- Landa, V., Shapira, Y., David, M., Karasik, A., Weiss, E., Reuveni, T., & Drori, E. (2021) Accurate classification of fresh and charred grape seeds to the varietal level, using machine learning based classification method. *Scientific Reports*, 11, 1-15. DOI: <https://doi.org/10.1038/s41598-021-92559-4>
- Machado, T. F. F., Andrade, A. M., Silva, A. P., Sena, M. F. M., & Thode-Filho, S. (2014). Potential energy of the woods leucena (*Leucaena leucocephala*) and pau-jacaré (*Piptadenia gonoacantha*). *Revista Eletrônica em Gestão, Educação e Tecnologia Ambiental*, 18(1), 45-50. DOI: <https://doi.org/10.5902/2236117013809>
- Matsumoto, N., Ogawa, M., Takayasu, K., Hirayama, M., Miura, T., Shiozawa, K., ... Udagawa, S. (2015). Quantitative sonographic image analysis for hepatic nodules: a pilot study. *Journal of Medical Ultrasonics*, 42(4), 505-512. DOI: <https://doi.org/10.1007/s10396-015-0627-3>
- Mazur, M., Marcysiak, K., Dunajska, A., Gawlak, M., & Kałuski, T. (2022). Taxonomic significance of seed morphology in *Veronica* L. (Plantaginaceae) species from central Europe. *Plants*, 11(88), 1-17. DOI: <https://doi.org/10.3390/plants11010088>
- Medina, W., Skurtys, O., & Aguilera, J. M. (2010). Study on image analysis application for identification quinoa seeds (*Chenopodium quinoa* Willd) geographical provenance. *Food Science and Technology*, 43(2), 238-246. DOI: <https://doi.org/10.1016/j.lwt.2009.07.010>
- Melito, S., La Bella, S., Martinelli, F., Cammalleri, I., Tuttolomondo, T., Leto, C., ... Mulas, M. (2016). Morphological, chemical, and genetic diversity of wild myrtle (*Myrtus communis* L.) populations in Sicily. *Turkish Journal of Agriculture and Forestry*, 40(2), 249-261. DOI: <https://doi.org/10.3906/tar-1505-118>
- Menegatti, R. D., Mantovani, A., Navroski, M. C., & Souza, A. G. (2017). Genetic divergence among provenances of *Mimosa scabrella* Benth. based on seed analysis. *Revista Brasileira de Ciências Agrárias*, 12(3), 366-371. DOI: <https://doi.org/10.5039/agraria.v12i3a5449>
- Metougui, M. L., Mokhtari, M., Maughan, P. J., Jellen, E. N., & Benhabib, O. (2017). Morphological variability, heritability, and correlation studies within an argan tree population (*Argania spinosa* (L.) Skeels) preserved *in situ*. *International Journal of Agriculture and Forestry*, 7(2), 42-51. DOI: <https://doi.org/10.5923/j.ijaf.20170702.02>
- Noronha, B. G., Pereira, M. D., Flores, A. V., Demartelaere, A. C. F., & Medeiros, A. D. (2018). Morphometry and physiological quality of *Moringa oleifera* seeds in the function of their fruit position. *Journal of Experimental Agriculture International*, 25(6), 1-10. DOI: <https://doi.org/10.9734/JEAI/2018/43375>
- Oliveira, S. S. C., Araújo-Neto, J. C., Cruz, S. J. S., & Ferreira, V. M. (2012). Morphometric characterization of seeds and plantules and *Anadenanthera macrocarpa* (Benth.) Brenan germination. *Ciência Florestal*, 22(3), 643-653. DOI: <https://doi.org/10.5902/198050986630>

- Orrù, M., Grillo, O., Lovi, G., Venora, G., & Bacchetta, G. (2012). Morphological characterisation of *Vitis vinifera* L. seeds by image analysis and comparison with archaeological remains. *Vegetation History and Archaeobotany*, 22(231), 231-242. DOI: <https://doi.org/10.1007/s00334-012-0362-2>
- Pan, X., Zhao, X., Yan, W., Liu, J., Luo, X., & Wuyun, T. (2021). *Identification of gramineous grass seeds using Gabor and locality preserving projections*. Beijing, CH: Springer Nature. DOI: https://doi.org/10.1007/978-981-16-3501-4_3
- Pontes, M. S., Montefusco-Pereira, C. V., Misra, B. B., Ribeiro-Junior, H. L., Graciano, D. E., Santos, J. S., ... Santiago, E. F. (2018). High-throughput phenotyping by applying digital morphometrics and fluorescence induction curves in seeds to identifying variations: a case study of *Annona* (Annonaceae) species. *Information Processing in Agriculture*, 5(4), 443-455. DOI: <https://doi.org/10.1016/j.inpa.2018.07.001>
- Rewicz, A., Bomanowska, A., Magda, J., & Rewicz, T. (2016). Morphological variability of *Consolida regalis* seeds of south-eastern and central Europe. *Systematics and Biodiversity*, 15(1), 25-34. DOI: <https://doi.org/10.1080/14772000.2016.1216017>
- Ropelewska, E., & Piecko, J. (2022) Discrimination of tomato seeds belonging to different cultivars using machine learning. *European Food Research and Technology*, 248, 685-705. DOI: <https://doi.org/10.1007/s00217-021-03920-w>
- Rosa, T. L. M., Araujo, C. P., Kamke, C., Ferreira, A., Ferreira, M. F. S., Oliveira, J. B., ... Alexandre, R. S. (2020). Sapucaia nut: morphophysiology, minerals content, methodological validation in image analysis, phenotypic and molecular diversity in *Lecythis pisonis* Cambess. *Food Research International*, 137, 1-9. DOI: <https://doi.org/10.1016/j.foodres.2020.109383>
- Roveri-Neto, A., & Paula, R.C. (2017). Variability among mother trees of *Ceiba speciosa* St. Hil for characteristics of the fruits and seeds. *Revista Ciência Agronômica*, 48(2), 318-327. DOI: <https://doi.org/10.5935/1806-6690.20170037>
- Santana, S. H., Torres, S. B., & Benedito, C. P. (2013). Biometrics of fruits and seeds and germination of balsam pear. *Revista Brasileira de Plantas Mediciniais*, 15(2), 169-175. DOI: <https://doi.org/10.1590/S1516-05722013000200001>
- Santos, J. C. C., Silva, D. M. R., Costa, R. N., Silva, C. H., Santos, W. D. S., Moura, F. D. B. P., & Silva, J. V. (2018). Biometric and morphological aspects of the fruits and seeds of *Schinopsis brasiliensis*. *Nativa*, 6(3), 219-224. DOI: <https://doi.org/10.31413/nativa.v6i3.4709>
- Silva, R. M., Cardoso, A. D., Dutra, F. V., & Morais, O. M. (2017). Biometric aspects of fruit and seed of *Caesalpinia ferrea* Mart. ex Tul. from semiarid baiano. *Revista de Agricultura Neotropical*, 4(3), 85-91. DOI: <https://doi.org/10.32404/rean.v4i3.1427>
- Silva, L. L. H., Oliveira, E., Calegari, L., Pimenta, M. A. C., Pimenta, A. S., & Dantas, M. K. L. (2018). Energy characteristics of charcoal aroeira (*Myracrodruon urundeuva* Allemão) and leucaena (*Leucaena leucocephala* (Lam.) R. de Wit). *Ciência Florestal*, 28(1), 412-419. DOI: <https://doi.org/10.5902/1980509831619>
- Tan, K., Wang, R., Li, M., & Gong, Z. (2019). Discriminating soybean seed varieties using hyperspectral imaging and machine learning. *Journal of Computational Methods in Sciences and Engineering*, 19(4), 1001-1015. DOI: <https://doi.org/10.3233/JCM-193562>
- Ucchesu, M., Orrù, M., Grillo, O., Venora, G., Paglietti, G., Ardu, A., & Bacchetta, G. (2016). Predictive method for correct identification of archaeological charred grape seeds: support for advances in knowledge of grape domestication Process. *PLoS ONE*, 11(2), 1-18. DOI: <https://doi.org/10.1371/journal.pone.0149814>
- Varma, V. S., Durga, K., & Keshavulu, K. (2013). Seed image analysis: its applications in seed science research. *International Research Journal of Agricultural Sciences*, 1(2), 30-36.
- Walker, K. P. (2012). Fodder potential of leaves and pods of planted *Leucaena diversifolia* and *L. leucocephala* species in semi-arid Botswana. *International Research Journal of Agricultural Science and Soil Science*, 2(10), 445-450.
- Wan, J. Z., & Wang, C. J. (2018). Expansion risk of invasive plants in regions of high plant diversity: A global assessment using 36 species. *Ecological Informatics*, 46, 8-18. DOI: <https://doi.org/10.1016/j.ecoinf.2018.04.004>

Wang, X. Y., Shen, D. W., Jiao, J., Xu, N. N., Yu, S., Zhou, X. F., ... Chen, X. Y. (2012). Genotypic diversity enhances invasive ability of *Spartina alterniflora*. *Molecular Ecology*, 21(10), 2542-2551.

DOI: <https://doi.org/10.1111/j.1365-294X.2012.05531.x>

Zuffo, A. M. A., Steiner, F., Alves, C. Z., Alcântara-Neto, F., Santos, M. D. A., Oliveira, A.M., ... Fonseca, W. L. (2019). Biometric characteristics of fruits, seeds and plants of *Hancornia speciosa* Gomes.

(Apocynaceae). *Australian Journal of Crop Science*, 13(4), 622-627.

DOI: <https://doi.org/10.21475/ajcs.19.13.04.p1651>