Fruit and seed biometry of *Carpotroche brasiliensis* (RB) A. Gray (Achariaceae), a tropical tree with great potential to provide natural forest products

Thâmara M. Lima¹*, Elizabeth S. Amaral¹, Fernanda A. Gaiotto¹, Leticia dos Anjos², Ândrea Carla Dalmolin³, Alessandro Souza Santos¹, Marcelo S. Mielke¹

¹Departamento de Ciências Biológicas, Universidade Estadual de Santa Cruz, Rodovia Jorge Amado km 16, Ilhéus, BA, Brazil  
²Departamento de Biologia, Universidade Federal de Lavras, Av. Central, s/n - Campus Universitário, Lavras, MG, Brazil  
³Centro de Formação em Ciências e Tecnologias Agroforestais, Universidade Federal do Sul da Bahia Br 415 Km 22 Ceplac, Ilhéus, BA, Brazil

*Corresponding author: thamaralima6@gmail.com

**Abstract**

*Carpotroche brasiliensis* (RB) A. Gray (Achariaceae) is a native tree species of the Brazilian Atlantic Forest with great potential to provide natural forest products in agroforestry systems. The oil of its seeds contains medicinal and cosmetic properties, and the fruits are appreciated by wild animals. In this study, we analyzed the biometry of fruits and seeds collected from naturally-grown trees in agroforestry systems on small farms in southern Bahia, Brazil. Fresh fruit mass (FFM), fruit length (FL), fruit diameter (FD), number of seeds per fruit (NSF), total fresh seed mass per fruit (FMSF), total dry mass of seeds per fruit (DMSF), and mass of 1,000 seeds (MTS) were collected from 66 fruits of 18 trees on six rural properties. Seed length (SL) and seed diameter (SD) were also evaluated for 5335 seeds. From this sample universe, 697 seeds represented an adequate sample size to measure these dimensions with statistical precision. The fruits analyzed in this study had uniform values for most of the biometric variables among the sites. The average values of NSF, FFM, and MTS were approximately 88 seeds, 0.5 kg, and 1.3 kg, respectively. Seeds of *C. brasiliensis* obtained from freshly-harvested fruits had high water content, with an average of more than 45%. Strongly significant correlations between FFM and FD and between FMSF and DMSF indicate the possibility of developing simple procedures to estimate seed production for commercial purposes from field evaluations.

**Keywords:** Indigenous trees, Phenotypic variability, Seed Sample Size, Network of correlation, Brazilian Atlantic Rainforest, Agroforestry systems.

**Introduction**

Tropical regions concentrate immense biodiversity in forest environments, conditioning the global importance of the tropics (Barlow et al., 2018). The southern coast of the state of Bahia, Brazil is recognized for its high diversity of tree species and for containing the highest percentage of tropical forest remnants of the Atlantic Forest in Northeast Brazil (Martini et al., 2007; Piotto et al., 2009). Today, southern Bahia is characterized by a mosaic landscape consisting of forest remnants and a rustic agroforestry system known as Cabrucas (Sambuchi et al., 2012). Cabrucas are an agroforestry system based on cacao (*Theobroma cacao*) cultivation under the shade of the tree strata of native forest remnants (Alvim et al., 1986; Johns, 1998). This traditional agroforestry practice has produced significant economic returns for the cocoa producers of the region (Piasentin et al., 2014). In addition to its socioeconomic importance, this system is a habitat for several species of local fauna and flora, and connects forest remnants (Cassano et al., 2011, 2016; Schroth et al., 2011). Currently, both Cabrucas and associated forest areas are threatened by the cocoa crisis (Schroth et al., 2011; Sambuchi et al., 2012), and there is a need for natural forest products that could potentially add value to the cocoa crop in multispecies agroforestry systems (Santos et al., 2012; Feijó et al., 2009). The management and exploitation of native trees for natural forest products in agroforestry systems can promote biodiversity conservation and farmers' livelihoods (Rao et al., 2004; Cerda et al., 2014). *Carpotroche brasiliensis* (RB) A. Gray (Achariaceae) is a tree species native to the Brazilian Atlantic Forest (Schulz et al., 1994; Marquete et al., 2015). A recent study showed that this species can be grown in agroforestry systems under the shade of taller trees as well as in forest enrichment programs (Cerqueira et al., 2017). In southern Bahia, *C. brasiliensis* is frequently found in forest remnants and cabbage, and has great potential for cultivation as a natural forest products, adding value to cocoa production (Schulz et al., 1994). It is a dioecious species, with male and female flowers in different individuals. The fruits are green with a fleshy orange-yellow pulp, which is greatly appreciated by wild animals, mainly rodents.
Souza et al. (2010). The seeds are numerous, polyhedral, with fine tests and a fleshy endosperm immersed in the pulp (Souza-Araújo, 1935). Chaumooogra oil is extracted from the seeds, and its constituents exhibit pharmacological activities used in the treatment of leprosy and skin diseases (Souza-Araújo, 1935). The oil extracted from the seeds also has anti-inflammatory, analgesic, and anti-parasitic properties (Lima et al., 2003; Oliveira et al., 2009). The prospective income from the cultivation of *C. brasiliensis* in multispecies agroforestry systems in southern Bahia is promising (Brito-Rocha et al., 2017; Cerqueira et al., 2017). Small farmers in the region sell the fruits of this species to cosmetics companies, which use the oil in manufacturing makeup products. Despite the commercial value of the tree, its exploitation is limited to extractivism, and studies of the species are still scarce, which represents an obstacle to the production of its fruits and seeds on a larger scale. Knowledge of the biology of native species is critical for the development of *in situ* and *ex situ* conservation programs and for the development of technologies for rational agroforestry systems management (Leakey et al., 1998). For example, the fruit and seed biometry of native trees can provide important information for studies of seed ecology (Kelly, 1995; Khurana et al., 2006) and seedling establishment (Baraloto et al., 2005), and can contribute to methods of seedling production (Sautu et al., 2006; Clifton-Cardoso et al., 2008; Macedo et al., 2009). Biometric data can also be used to estimate fruit and seed production. In the context of domestication, knowledge of production can contribute to selecting beneficial genotypes and improving agronomic attributes in order to develop plantations with desirable characteristics for commercialization (Gusmão et al., 2006; Leão et al., 2018; Zuffo et al., 2019). Because it contributes to the description of phenotypic variation, biometric characterization provides vital information for the conservation and exploitation of genetic resources (Schwartz et al., 2007; Silva et al., 2017).

Given these circumstances, and considering the natural occurrence of *C. brasiliensis* in agroforestry sites in southern Bahia, the biometric characterization of its fruits and seeds will allow the evaluation of the performance of this species in the region (Botellelli et al., 2000; Menegatti et al., 2017). Determining the biometric variables of fruits and seeds can be a way of identifying useful linear relationships between variables (Maurya et al., 2015; Zuffo et al., 2016). As the seeds are of interest for propagation and are the main commercial product extracted from this tree, another potential advantage of this type of characterization is the possibility of estimating, based on statistical parameters, the number of seeds required to measure seed dimensions with statistical accuracy (Cargnelutti Filho et al., 2012; Schabarum et al., 2018). Due to the fact that many of these characteristics are relatively unknown for this species, the general objective of this study was to describe the biometry of fruits and seeds collected from plants grown naturally in agroforestry systems in small rural properties in southern Bahia, Brazil. Based on this approach, we had the following specific objectives: (i) to investigate the variation in biometric characteristics among agroforestry sites, (ii) to analyze the correlations between these biometric characteristics, and (iii) to evaluate the influence of sample size on the accuracy of seed dimension estimations.

**Results**

**Description of the biometry of *C. brasiliensis* fruits and seeds in the study area**

In general, fruit and seed biometry were quite similar among sites (Fig. 2). The average fruit mass was approximately 0.5 kg. The lowest and the highest mean FFM values were found in S4 (400.8 g) and S2 (617.4 g), respectively (Fig. 2A). The lowest and highest individual values for FFM were found in S5 (114.2 g) and S2 (1065.4 g), respectively. For all sites, FL (Fig. 2B) was slightly greater than FD (Fig. 2C). The average values for FL and FD across all sites were 12.2 cm and 10.7 cm, respectively. The lowest values for FL and FD were also observed in S4. Number of seeds per fruit (NSF) ranged from 71.8 (S4) to 103.3 (S3) (Fig. 2D), and the average NSF value was 88 seeds. The average FMSF and DMSF values ranged from 95.4 g (S1) to 127.0 g (S3) (Fig. 2E), and from 49.1 g (S1) to 66.9 g (S2) (Fig. 2F), respectively. The total average values for SL (Fig. 2G) and SD (Fig. 2H) were 14.1 mm and 11.8 mm, respectively. The average MTS values ranged from 1,023.5 g (S1) to 1,576.1 g (S2) (Fig. 2I), with a mean value of 1,313.4 g (N = 76 fruits). This result indicates that every 1.3 kg of fresh *C. brasiliensis* seeds contains around 1,000 seeds in southern Bahia. The mean SWC (Seed water content) value was approximately 46% across all sites. Our results indicate that fruits harvested from trees in southern Bahia have a fresh weight greater than 0.8 kg, and lengths and diameters greater than 15 cm and 10 cm, respectively.

**Similarity of the sample sites according to the set of fruit and seed biometric variables.**

The cluster trend measure calculated using the Hopkins statistic (H) showed the viability of the cluster analysis for our dataset (H = 0.50). According to the analysis of group number, the dissimilarity among sites was equal to one (Fig. 3A). In this sense, it was possible to visualize a single group among the agroforestry sites using the K-means grouping method, according to the biometric similarities of their fruits and seeds (Fig. 3B). The non-significant variation among the agroforestry sites, verified by multivariate analysis, was consistent with the results of the descriptive statistics for each biometric variable, which indicated similarity among the sites.

**Correlation network of biometric variables of fruits and seeds**

The network shown in Figure 4 shows the correlations between the fruit and seed biometric variables. Purple nodes represent seed variables, while yellow nodes represent fruit variables. The constructed network shows the grouping of these two types of variables. According to the thickness of the network edges, the magnitude of the correlation coefficient mainly varied for FL-SD and for FMSF- DMSF, DMSF-FD, FD-FFM, and FFM-FMSF. Of the strongly positive correlations, significant correlations were observed for FMSF-DMSF, FFM-FD, and NSF-FMSF. A significant negative correlation was verified for NSF-FFM. The MTS, SL, and SD variables did not present significant correlations with any other variables.
Fig 1. Map of the study area in the south of Bahia, Brazil, highlighting areas without forest (Non-Forest) and areas with forest remnants (Forest); Spatial distribution of the sampled *Cabrucu* agroforestry sites (S1, S2, S3, S4, S5, S6).

Fig 2. Fruit fresh mass (FFM) (a), fruit length (FL) (b), fruit diameter (FD) (c), number of seeds per fruit (NSF) (d), total fresh mass of seeds per fruit (FMSF) (e), dry mass of seeds per fruit (DMSF) (f), mass of 1,000 seeds (MTS) (g), seed length (SL) (h), and seed diameter (SD) (i) of *C. brasiliensis* trees from six sites in southern Bahia, Brazil. (n = three trees per site).

Fig 3. Number of significant groups defined for grouping (a); Clustering of agroforestry sites (S1, S2, S3, S4, S5, S6) as a function of all the biometric characteristics of *C. brasiliensis* fruits and seeds (b).
Fig 4. Pearson’s correlations between network biometric variables measured in *C. brasiliensis* fruits at six sites in southern Bahia, Brazil. Green and red edges correspond to positive and negative correlations, respectively. The width and intensity of the edges indicate the absolute value of the correlations. Fruit fresh mass (FFM), fruit length (FL), fruit diameter (FD), number of seeds per fruit (NSF), total fresh mass of seeds per fruit (FMSF), dry mass of seeds per fruit (DMSF), mass of 1,000 seeds (MTS), seed length (SL), and seed diameter (SD) (n = 76 fruits).

Fig 5. Sample sizes needed to minimize the coefficient of variation (a); margin of error (b).

**Seed sample size**

Based on the length and diameter data of 5335 seeds, an analysis that relates sample size (number of seeds) and coefficient of variation revealed that our estimates of the dimensions had good statistical accuracy. The results show that it would be necessary to measure only 697 seeds to obtain a coefficient of variation lower than 5% (Fig. 5A). To reach a relative error rate of 3%, measuring at least 3284 seeds would be sufficient (Fig. 5B).

**Discussion**

Given that systematic information on the biometry of *C. brasiliensis* fruits and seeds is practically non-existent, our study contributes to the enlargement of our biological knowledge of this species. To our knowledge, apart from a field survey conducted in the early twentieth century by Souza-Araujo (1935), no other research effort has been made on the biometric characterization of *C. brasiliensis* fruit and seeds to date.

The mean values of FFM, FL, and FD for all fruits in this study agree with those reported by Souza-Araujo (1935) in other regions of Brazil. Information on the mass of 1,000 seeds, such as those obtained in our research, are important for calculating the number of seeds necessary to plant in forest nurseries and in the industrial use of seeds. Foster (1982) proposed a classification system for seeds of tropical tree species based on six seed mass size-cATEGORIES. According to average values obtained for MTS in this study, the fruits of *C. brasiliensis* from southern Bahia contain 634.5 to 977 seeds kg\(^{-1}\), which includes this species in category five of large seeds (100-1000 seeds kg\(^{-1}\)) of Foster’s classification. Sautu et al. (2006), using the methodology proposed by Foster (1982), conducted a study on seeds of 99 rainforest tree species in Panama. They found that only 16.2% of the species studied could be included in category five, and that 80% of the species were included in categories one to four (>1,000 seeds kg\(^{-1}\)). Considering the data from Sautu et al. (2006) as a comparison, we could infer that *C. brasiliensis* is within a particularly small percentage of tropical species that have large seeds. Research results of this nature represent a reliable metric for predicting the survival and growth of seedlings, especially in forest restoration areas (Baskin et al., 1998). In a tropical forest scenario with frequent and negative impacts of anthropic interventions, biological
knowledge of tropical tree species becomes important for supporting restoration projects, particularly in the Atlantic Forest, which is a threatened tropical hotspot (Martini et al., 2007; Ribeiro et al., 2009; Barlow et al., 2016; Macera et al., 2017).

The SWC values in found this study can be considered high when compared with other tropical tree species (Walters, 2000; Sautu et al., 2006; Clifton-Cardoso et al., 2008), which may indicate that the seeds of this species are recalcitrant (Walters, 2000). Seed water content (SWC) values of around 40-50% have been reported for other tropical rainforest tree species that lose their germination potential after desiccation, even when seeds have high SWC values (Clifton-Cardoso et al., 2008). Further studies are needed to investigate whether seeds of *C. brasiliensis* are recalcitrant. However, in natural forests, or when the species is grown in agroforestry systems, it is common to find broken fruits on the ground containing large number of seedlings and germinating seeds. In addition, preliminary tests made in small nurseries in southern Bahia have shown loss of seed viability after short storage periods (weeks or a few months).

Taking the results altogether, we can suggest that fruits and seeds of *C. brasiliensis* grown naturally in southern Bahia are uniform in biometry among populations in different agroforestry sites. Cabruca is an agroforestry system with unique characteristics due to its level of biodiversity of fauna and flora. These peculiarities promote important agroecological services for species productivity and stability (Johns 1998; Sambuichi et al., 2007; Schroth et al., 2007). The observed uniformity in species-seed biometry may be linked to the ability of this agroforestry system to support pollination, soil fertility, nutrient cycling, and biological control of pests and diseases (Young, 1982; Moço et al., 2009; Fontes et al., 2014; Piasentin and Saito 2014; Aleixo et al., 2017).

Correlations between fruits and seeds can help us to understand the dynamics of the inherent relationships between the characteristics of each species (Gonçalves et al., 2013; Souza et al., 2016; Mauyra et al., 2015; Zuffo et al., 2016). The network correlation analysis performed in our study allowed the visualization of correlations and binding patterns that are difficult to detect in numerical correlation matrices (Epskamp et al., 2012). This knowledge is vital for the domestication of wild species, as is the case in our biological model of study (Silva et al., 2016). Significantly strong correlations between FFM-FD and FMSF-DMSF indicate that it is possible to develop procedures to estimate seed production for commercial purposes from field evaluations. The strategy of selecting more productive genotypes through the characteristics of each species (Gonçalves et al., 2013; Souza et al., 2016; Mauyra et al., 2015; Zuffo et al., 2016). The network correlation analysis performed in our study allowed the visualization of correlations and binding patterns that are difficult to detect in numerical correlation matrices (Epskamp et al., 2012).

Considering that the potential harvest from *C. brasiliensis* in agroforestry systems is mainly the seeds, which also represent the propagation structure, our study provided knowledge that may help to subsidize the efficiency of the use of this species. In addition to what has already been discussed regarding the seeds, another important approach is the ideal sample size in the context of biometric characterization. For native trees such as *C. brasiliensis*, irregular seed production may hinder robust sampling. The greater the sample size used to estimate variables, the lower the coefficient of variation and, consequently, the lower the error estimate (Sears, 1964; Kelley, 2007). The estimated seed size in our study was established through a large sample size (5335 seeds measured). As a smaller number of seeds than that evaluated would already be sufficient to obtain good precision, adjusting the sample size according to statistical parameters contributes to the efficiency of robust results (Schabaran et al., 2018), mainly for native trees.

Materials and methods

**Plant materials**

The fruits were collected from spontaneously-growing trees in six small-farm sites (S1-6) following the Cabruca agroforest system in the municipalities of Camamú and Maruá, Bahia, Brazil. A heterogeneous landscape mosaic map of the study area was created using ArcGis (10.2) (Figure 1). The fruits were collected between September and November 2013.

**Traits measured**

Fruit fresh mass (FFM), fruit length (FL), fruit diameter (FD), number of seeds per fruit (NSF), total fresh mass of seeds per fruit (FMSF), total dry mass of seeds per fruit (DMSF), and mass of 1,000 seeds (MTS) were measured from 66 fruits of 18 trees. The protocols of the International Seed Testing Association (Ista, 1993) and the rules of the Brazilian Ministry of Agriculture and Food Supplies for seed analysis (Mapa, 2009) recommend that MTS should be determined for eight replicates of 100 seeds. In our study, MTS was calculated from the average of 76 fruits with approximately 88 seeds each. In total, we evaluated 5335 seeds, which provided a good measure of the mass of 1,000 seeds for the species in southern Bahia. Seed length (SL) and seed diameter (SD) were evaluated using a digital caliper. Total dry mass of seeds per fruit (DMSF) was obtained after drying the leaves at 75°C until a constant mass was reached. Mass of 1,000 seeds (MTS) was calculated using the following equation: MTS = (1000 x FMSF)/NSF. The water content of the seeds (SWC) was calculated as $SWC = \frac{FMSF - DMSF}{FMSF}$. The absolute value of the correlation between these nodes is represented by boxplot graphs. Pearson correlation was estimated between the variables studied through a correlation network. The correlation network analysis was based on the Fruchterman-Reingold algorithm (Fruchterman et al., 1991), which shows the proximity between nodes by the absolute value of the correlation between these nodes (Epskamp et al., 2012). Each node represents a variable, and each edge represents a correlation between two variables. The ‘igraph’ package (Epskamp et al., 2012) in R was used to carry out the network analysis. The significance of the Pearson correlation coefficients was estimated by Student’s t-tests at a 5% probability level. To analyze the similarities between the sites (S1-S6), a multivariate analysis was performed using the K-means non-hierarchical clustering method (Macqueen, 1967). It is noteworthy that before choosing a clustering approach, Hopkins statistic (H) was performed to evaluate the clustering trend (Banerjee et al., 2004). The analysis was performed based on the number of groups defined by the clusgap function and on clustering by the hmeans function of the ‘cluster’ and ‘factoextra’ packages in R, respectively (Tibshirani et al., 2001). In addition, the ‘samples4’ package was used to estimate the appropriate
sample size according to the coefficient of variation (Gutiérrez et al., 2016). All statistical analyses were completing in the R version 3.5.3 programming software (R Development Core Team, 2019).

Conclusions
This study provided the biometric characterization of C. brasiliensis fruits and seeds in southern Bahia, Brazil. Relatively unknown data on the biometric attributes of this species indicate the importance of our study in developing strategies for the seeds of the species to be used as a potential natural forest products. The lack of variation among agroforestry sites indicates a process of incipient domestication of this tree. The fact that the oil extracted from C. brasiliensis seeds is valuable as a commercial product probably did not allow selection intensity to affect variation among sites. Our study also estimated the number of seeds suitable for biometric characterization. We emphasize the importance of this particular analysis in establishing a suitable degree of statistical confidence in research on forest species that usually produce a smaller seed sample size than agricultural species.

Acknowledgments
This study was supported by Natura Inovação e Tecnologia de Produtos LTDA. We thank Gerson J. Sales Neto, Nilson A. dos Santos, and Rones F. Souza, of Floresta Viva Institute, and Meide Fernando B. Santana for assistance with data collection. We also acknowledge the farmers of the municipalities of Camamu and Marãs, southern Bahia, Brazil, for consent the collection of fruits in their properties. FAG and MSM gratefully acknowledge productivity fellowship provided by CNPq (Brazilian National Council for Scientific and Technological Development). TML acknowledges CAPES for her Ph.D. scholarship.

References


