



## Article

# An Angiosperm Species Dataset Reveals Relationships between Seed Size and Two-Dimensional Shape

Emilio Cervantes <sup>1,\*</sup> , José Javier Martín Gómez <sup>1</sup> , Diego Gutiérrez del Pozo <sup>2</sup> and Luís Silva Dias <sup>3</sup> 

<sup>1</sup> IRNASA-CSIC, Cordel de Merinas, 40, E-37008 Salamanca, Spain; jjavier.martin@irnasa.csic.es

<sup>2</sup> Department of Wild Life Conservation and Magamenent, Amazon State University (UEA), Napo EC-150950, Ecuador; dgutierrez@uea.edu.ec

<sup>3</sup> Department of Biology, University of Évora, Ap. 94, 7002-554 Évora, Portugal; lsdias@uevora.pt

\* Correspondence: emilio.cervantes@irnasa.csic.es; Tel.: +34-9-2321-9606

Received: 25 June 2019; Accepted: 27 September 2019; Published: 9 October 2019



**Abstract:** Datasets containing information on seed size have been published and are currently available. Nevertheless, there is a lack in the literature of a dataset dedicated to seed shape. We present a preliminary version for a dataset on seed morphology based on a comparison of seed shape with geometric figures. Similarity of the outline of seed images with geometric models is considered as a basis to classify seeds according to the geometric figures they resemble (e.g., ellipse, oval, cardioid). This allows, first, the classification of plant species according to their geometric type of seed, and second, seed shape quantification. For each seed image, the percent of similarity of their outline with a geometric figure can be calculated as a J index. Similarity in absolute terms is considered only when the J index >90. This criterion is important to avoid ambiguity and increase discrimination. The dataset opens the possibility of studying the relationship between seed shape and other variables such as seed size, genome complexity, life form or adaptive responses.

**Keywords:** geometric curves; J index; morphology; shape; seed

## 1. Introduction

### 1.1. Difficulties in the Description of Seed Shape

There is some confusion regarding seed shape in the botanical literature. First, a long-standing tradition keeps using epithets that describe shapes in a rather arbitrary form, such as kidney-shaped, reniform, meloniform, etc. These are of little utility because there is not a geometric object well defined such as the kidney or a melon. Second, the concept of shape is often confused with size. Third, examples are abundant where length-derived magnitudes are taken as measurements of seed shape. We will analyze some cases for each of these three sources of confusion.

#### 1.1.1. Terms Used to Describe Shape

Epithets such as kidney-shaped, reniform, meloniform [1], refer to objects (kidney, melon) that resemble geometric figures (the cardioid and the oval, for example), but are not themselves well described geometric figures. Similarly, Delorit and Gunn refer to some seeds as “mitten-shaped” [2] and other authors refer to the same seeds as “globose-ovoid” (*Trifolium arvense*) or “globose-lenticular” (*Trifolium subterraneum*) [3], or “obovate” (*Trifolium hybridum*), “ovate” (*Trifolium fragiferum*) or “elliptic” (*Trifolium dubium*) [4], but also other epithets apply to the same forms such as “large-cordate to asymmetric kidney-shaped” (*Crotalaria pallida*) or “cordate-kidney-shaped to globose-triangular,

ovate-orbicular in outline” (*Ulex europaeus*) [5]. While most of these descriptions are not very useful, descriptors such as ovoid, elliptic, piriform and heart-shaped are adequate because in geometry there are defined figures termed the oval [6], ellipse [7], piriform curve [8], and heart curve [9]. Only in these latter cases can the objects be well-defined and quantified by comparison.

### 1.1.2. Shape Is Often Confused with Size

A second frequent problem is that the concept of shape is often confused with size, thus making impossible further development of shape analysis. In a recent example [10], the authors described OsGRF4, a gene involved in panicle length and seed shattering in rice. The gene encodes a growth-regulating factor involved in the regulation of two cytokinin dehydrogenase precursor genes (CKX5 and CKX1). Alterations in this gene resulted in modifications of size but the authors confused seed size and shape when they stated, “Thus, the increase in grain weight and yield per plant resulted from increase in both grain shape and grain filling rate.” The expression, “increase in grain shape” is incorrect, because if there is not a magnitude defined for seed shape, then there cannot be an increase in seed shape. In the absence of a magnitude that may refer to seed shape, the reported increases are related to seed size and not shape.

Also, Hu et al. [11] used recombinant inbred lines (RILs) and cultivated soybean accessions to evaluate seed length, seed width and seed height as seed size traits, and the ratios of these values were described as seed shape traits. But these ratios, albeit often used in this context, are very poor evaluators for seed shape. Related to this, Tanabata et al. [12], when describing the software SmartGrain stated, “Outlines of seeds are automatically recognized from digital images, and several shape parameters, such as seed length, width, area, and perimeter length, are calculated.” However, length, width, area, and perimeter are very poor morphological descriptors. Figures having very similar measures in all these parameters may be very different in shape. Similarly, ratios of these magnitudes are also incomplete descriptors of shape.

### 1.1.3. Variations in Width and Height Are Poor Measures of Shape

Sometimes the estimation of shape has been based on the variation in the orthogonal dimensions of seeds. Seed length is considered as the reference and the ratios of width/length and height/length are measured from the basis that these ratios are equal to one for a sphere. The estimation of volume based on these measurements assumes that the seeds are ellipsoids, and this method indicates how the shape of an object deviates from an idealized sphere. Thus, Bekker et al. [13] stated, “The shape of seeds (Vs.) can be captured by measuring length, width and height of a seed and dividing all values through length so length is unity and then calculating the variance of these three values by dividing the summed squared deviation from the mean by  $n - 3$ :  $\sum (x - \bar{x})^2/n$ .” but this is not an accurate measure of the shape of an object. It is only a measure of how much a figure diverges from a sphere.

## 1.2. The Advantage of the Geometric Criterion J Index

A criterion must be established for description and classification of seed shape with the objective that the description is productive i.e., open to quantification and useful for classification. Comparing the shape of seeds (outline of seed images showing length and width) to geometric figures fulfills some conditions: (1) It permits a precise description and quantification of seed shape, and (2) shape description is independent of size. For this purpose, the epithets used to define the shapes of seeds must be based on geometric figures taken as models. This allows quantification of seed shape by comparison of the outline of the image of a seed with a corresponding geometric figure. The J index indicates the percent of similarity between both images, the seed and the geometric figure [14–23]. It ranges between 1 and 100 and is a measure of shape, not size.

## 1.3. Datasets on Seed Size: A Review

Size is a quantitative character and in relation to seeds it may be represented either as mass (grams) or more correctly as volume ( $\text{mm}^3$ ). Several datasets such as the Seed Volume Dataset (SVDS;

SVDS\_03plus.csv) contain data for seed size [24] and weight [25–32] for many plant species. The dataset ECE-2017-06-00864.R1-Dryad (ECE) [25] is based on the data from a Kew dataset of seeds (SID 2.0, January 2002) [26], and is related to other work [27–32].

Before developing a seed shape dataset, these two “seed size” datasets were examined. The SVDS [24] contained data for more than 6500 species and varieties without previous geographic selection, while the ECE [25] contained data for 1600 species and subspecies in the Pannonian Ecoregion in Central Europe. The distributions of size in both datasets were compared. Ordering species by increasing size, the results for both datasets were more similar for species with small seeds than for large seeds.

Among the 50 species with the smallest seeds in the ECE, 37 belonged to the Orchidaceae (74%). Other species belonged to the Orobanchaceae (4), Pyrolaceae (4), Monotropaceae (2), Scrophulariaceae (1), Asteraceae (1) and Gentianaceae (1). Among the 50 species with the smallest seeds in the SVDS, a majority also corresponded to the Orchidaceae (36 species, 72%), which is normal because Orchidaceae is the most diverse plant family in the world; and others to the Orobanchaceae (5), Scrophulariaceae (3), Caryophyllaceae (1), Cyperaceae (1), Elatinaceae (1), Fabaceae (1), Juncaceae (1) and Lythraceae (1).

However, the results differed more between both datasets concerning large seeds. The 50 species having the largest seeds in the ECE corresponded to 22 families: Alliaceae, Amaryllidaceae, Apiaceae, Apocynaceae, Asteraceae, Boraginaceae, Brassicaceae, Caryophyllaceae, Convallariaceae, Convolvulaceae, Fabaceae, Geraniaceae, Helleboraceae, Hyacinthaceae, Iridaceae, Liliaceae, Paeoniaceae, Poaceae, Rosaceae, Rubiaceae, Scrophulariaceae and Solanaceae. While the 50 species having the largest seeds in the SVDS belonged to 21 families: Anacardiaceae, Apocynaceae, Araucariaceae, Arecaceae, Betulaceae, Boraginaceae, Clusiaceae, Combretaceae, Ebenaceae, Euphorbiaceae, Fabaceae, Hippocastanaceae, Icacinaceae, Juglandaceae, Lamiaceae, Lecythidaceae, Malvaceae, Olacaceae, Rosaceae, Rutaceae and Sapotaceae.

Thus, preliminary conclusions were: (1) There is more taxonomic variation in the families concerning large seeds than small seeds, and (2) when comparing a local dataset with a global dataset, there is more difference between local and global floras in the largest seeds than in the smallest seeds (the SVDS concerns species of global geographic origin and the ECE is concentrated in the Pannonian Ecoregion in Central Europe). This is in agreement with the reported result that seed size decreases with increasing latitude from the equator [32]. Thus, plants in the local dataset contained species representative of world flora with respect to the smallest seeds (orchids and Orobanchaceae) but not in relation to the largest seeds.

The analysis of these sized datasets allows other, more general, conclusions. For example, although there are families with small seeds (Orchidaceae) and others with large seeds (Arecaceae), there is not a clear, general association with a size-range characteristic for each family. Many families contain simultaneously species with both small and large seeds. For example, in the Acanthaceae, the seeds of *Elytraria crenata* have a volume of  $0.015 \text{ mm}^3$  while those of *Acanthus spinosus* have a volume of  $326 \text{ mm}^3$ . In between these two sizes there are seeds of many plant families. Similarly, in the Anacardiaceae, *Schinus terebinthifolius* seeds have an estimated volume of about  $1 \text{ mm}^3$ , while the volume of *Mangifera indica* seeds is about  $176,295 \text{ mm}^3$ .

Among the species with large seeds, the quantity of species that are from the Pacific Islands is surprising, and also that they belong to a diverse number of families. For example, in the Hawaiian Islands there are species with seeds larger than  $1000 \text{ mm}^3$  in the Anacardiaceae, Apocynaceae, Arecaceae, Boraginaceae, Clusiaceae, Combretaceae, Ebenaceae, Euphorbiaceae, Fabaceae, Malvaceae, Meliaceae, Rosaceae, Rutaceae, Sapindaceae and Sapotaceae. Similarly, there are orders with small seeds (Saxifragales, Rubiales) and orders with large seeds (Fagales, Arecales, Malpighiales and Icaciniales), but nevertheless size is not a property adequate for the classification of the orders because in most of them we find species with large and small seeds.

The availability of a dataset of seed shape may help in investigating how the diverse types of seed shape are distributed in the taxonomic groups and how they correlate with size. The hypothesis is that

quantification of seed shape by comparison with geometric figures may help clarify the relationships between seed shape and size, as well as in the long term with plant lifestyle, and contribute to new approaches for taxonomy and phylogeny.

## 2. Materials and Methods

### 2.1. An Important Methodological Question

For this work, the concept of “seed” is viewed in a broad sense and it may include, in some instances, diaspores that correspond in a strict morphologic sense to fruits like achenes (in some species in the Ranunculaceae), caryopses (in the Poaceae), cypselae (in the Asteraceae), nut pericarps (in the Fagaceae and Betulaceae), mericarps (in the Malvaceae), drupes (in the Rosaceae) and others.

### 2.2. Dataset Construction

The dataset of seed shape was made with published results [14–23] or new measurements performed for this article as indicated (the dataset is provided in Supplementary Information; it contains a list of references for the entries). In general, the method was always the same: (1) Search for a geometric model that fits the contour of a given seed image, and (2) calculate the percent of similarity between seed image and geometric figure (J index). For the more common geometric figures (cardioid, ellipse, oval), only the images giving results over 90 were included in the dataset. This resulted in a reduced number of species included but the advantage was that the species represented were associated with a given geometric model with high accuracy.

In addition to the curved geometric models considered in the review presented in this issue (cardioid, ellipse, oval, heart-shaped curves, outline of Fibonacci’s spiral and lens [14]), the truncated pyramid was included, albeit there was no quantification by J index with this model due to shape variation in the seeds. Data on seed size (volume) and weight, taken respectively from [24] and the Kew dataset [26] were added to entries.

The relationship between seed size and shape was investigated combining the dataset of seed shape with the seed volume dataset (SVDS) [24]. For each species with geometrically defined shape in the former its presence in the latter was checked so that its volume could be determined. SVDS data still unpublished was also used and synonymy was checked [33]. The analysis involved only those species for which both data of morphology (J index > 90) and volume were known (60 species with cardioid seeds, 24 species with elliptic seeds, and 19 species with oval seeds). Species with Fibonacci’s spirals, lens, or truncated pyramid seeds were too few for meaningful analyses. The minimum seed volume was used throughout as a surrogate of seed size (rationale for this decision can be found in [24]). Size–number cumulative frequencies of species with cardioid, elliptic, or oval seeds were generated and investigated using the Weibull equation [34]:

$$CF = 1 - \exp \{- [ (SV - l)/k ]^c \}, \quad (1)$$

where  $CF$  is the cumulative relative frequency at seed volume,  $SV$ ;  $l$  is a location parameter that estimates the highest volume at which  $CF$  is strictly zero, which in practice represents  $SV$  of the smallest seed ( $SV_s$ ) thus equating  $l$  with  $SV_s$  which is a constant;  $k$  is a scale parameter with  $l + k$  estimating the volume corresponding to 63% of cumulative frequency;  $c$  is a dimensionless shape parameter estimating the symmetry of the size–number distribution, with  $3.25 \leq c \leq 3.61$  showing symmetry and representing a good approximation to the normal distribution,  $c < 3.25$  positive asymmetry,  $c > 3.61$  negative asymmetry [35,36]. Coefficients of determination of fitted Weibull equations were 0.994 (cardioid seeds), 0.977 (elliptic seeds), and 0.954 (oval seeds).

### 2.3. Weibull Equations

Weibull equations were fitted with Statgraphics Plus ver. 3.3 (Manugistics, Rockville, MD, USA) by least squares nonlinear regression without replication using the Marquardt method [37]. The analysis involved only those species for which both data of morphology (J index > 90) and volume were known, resulting in 60 species with cardioid seeds, 24 species with elliptic seeds, and 19 species with oval seeds.

### 2.4. Statistical Analysis

ANOVA was used to ascertain significant differences in volume between seed populations corresponding to each model, followed by Scheffé's test ( $p < 0.05$ ) to provide specific information on which means were significantly different from one another. The analysis was done with software IBM SPSS statistics v25.

## 3. Results

### 3.1. An Overview of the Dataset of Seed Shape

The dataset in its current status (Supplementary Information) contained a total of 239 species belonging to 63 families (27 orders), with their values of J index. The most frequent geometric models were the cardioid (134 species), oval (46 species), and ellipses of varied x/y ratios (49 species). The oval was present in species of the following 17 families (given in alphabetic order by name of the corresponding orders [11 orders]).

Asparagaceae (Asparagales); Asteraceae (Asterales); Schisandraceae (Austrobaileyales); Boraginaceae (Boraginales); Akaniaceae, Cleomaceae (Brassicales); Cucurbitaceae (Cucurbitales); Adoxaceae (Dipsacales); Cyperaceae (Poales); Lardizabalaceae, Ranunculaceae, Berberidaceae, Euptelaceae (Ranunculales); Rhamnaceae, Rosaceae, Cannabaceae (Rosales) and Crassulaceae (Saxifragales).

Diverse ellipses with varied x/y ratios were observed in 49 species of the 23 families belonging to 17 orders: Araceae and Hydrocharitaceae (Alismatales); Amborellaceae (Amborellales); Areaceae (Arecales); Asparagaceae (Asparagales); Calyceraceae, Campanulaceae, Menyanthaceae (Asterales); Akaniaceae and Cleomaceae (Brassicales); Celastraceae (Celastrales); Anisophylleaceae, Begoniaceae, Corinocarpaceae and Cucurbitaceae (Cucurbitales); Fabaceae (Fabales); Fagaceae (Fagales); Apocynaceae (Gentianales); Euphorbiaceae (Malpighiales); Poaceae (Poales); Ranunculaceae (Ranunculales) and Elaeagnaceae and Rosaceae (Rosales).

The cardioid or modified cardioid was present in 134 species belonging to 28 families in 10 orders: Brassicaceae, Capparaceae, Cleomaceae and Resedaceae (Brassicales); Achatocarpaceae, Aizoaceae, Amaranthaceae, Basellaceae, Cactaceae, Caryophyllaceae, Chenopodiaceae, Gisekiaceae, Lophiocarpaceae, Molluginaceae, Petiveriaceae, Physenaceae, Phytolaccaceae, Portulacaceae and Talinaceae (Caryophyllales); Primulaceae (Ericales); Fabaceae (Fabales); Malvaceae (Malvales); Elaeocarpaceae (Oxalidales); Papaveraceae (Ranunculales); Anacardiaceae and Sapindaceae (Sapindales); Solanaceae (Solanales) and Lowiaceae (Zingiberales).

Other figures were less represented, such as the heart-shaped curves, outline of Fibonacci's spiral, and lens. Heart-shaped curves were typical of the Vitaceae (Vitales). Seeds whose profile resembled the outline of a Fibonacci's spiral were found in the Alismataceae (Alismatales), Coriariaceae (Coriariales) and also in some examples of the Ranunculaceae (Ranunculales). The lens shape was observed in the Hydrocharitaceae and Posidoniaceae (Alismatales), Amborellaceae (Amborellales) Celastraceae (Celastrales), Strasburgeriaceae (Crossosomatales), Cucurbitaceae (Cucurbitales) as well as some species and varieties of wheat (*Triticum monococcum*, einkorn) [38].

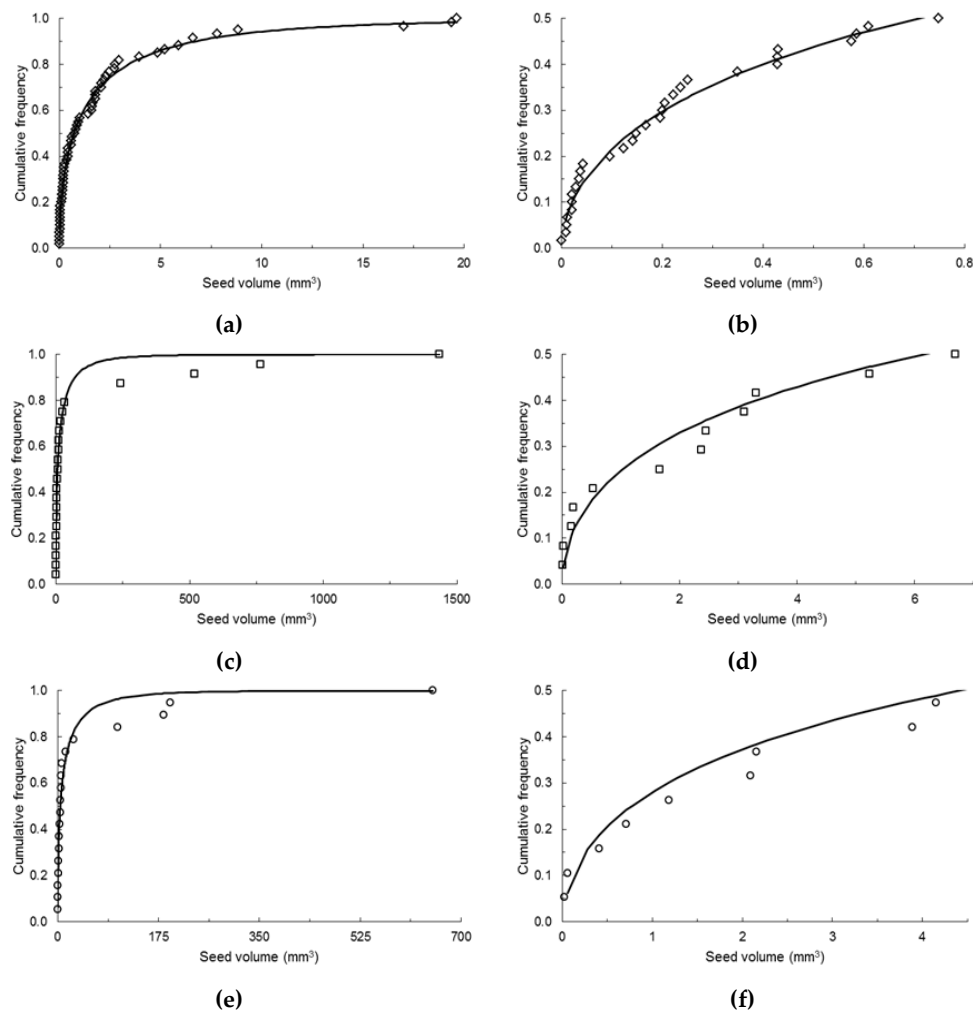
### 3.2. Relationship between Seed Size and Shape

In general, large seeds did not have the same shapes as small seeds. Elongated forms were more predominant in small seeds whereas "rounded" shapes tend to be more frequent in large seeds.

The relationship between seed size and seed sphericity is a polemic one. There are reports indicating a negative relationship between weight and sphericity of diaspores, more evident in the case of seeds, less so in fruits [39]. Size of cardioid seeds ranged from  $4.72 \times 10^{-4}$  to  $19.63 \text{ mm}^3$  (order of magnitude from  $-4$  to  $1$ ), with an observed median of  $0.748 \text{ mm}^3$  (Table 1, Figure 1a,b). Size of elliptic seeds ranged from  $0.007$  to  $1434.50 \text{ mm}^3$  (order of magnitude from  $-3$  to  $3$ ), with an observed median of  $6.691 \text{ mm}^3$  (Figure 1c,d). Size of oval seeds ranged from  $0.021$  to  $651.49 \text{ mm}^3$  (order of magnitude from  $-2$  to  $2$ ), with an estimated median of  $4.417 \text{ mm}^3$  (Figure 1e,f). ANOVA revealed differences between samples and the result of Scheffé test is shown in Table 2.

**Table 1.** Summary of data corresponding to seed volume ( $\text{mm}^3$ ) for each of the three geometric models used.

| Shape    | Number of Species | Mean Volume | Standard Deviation | Standard Error | Minimum Volume (mm)   | Maximum Volume (mm) | Volume Range (mm) |
|----------|-------------------|-------------|--------------------|----------------|-----------------------|---------------------|-------------------|
| Cardioid | 60                | 2.4         | 4.3                | 0.5            | $4.72 \times 10^{-4}$ | 19.6                | 19.6              |
| Ellipse  | 24                | 139.2       | 334.1              | 68.2           | $7.15 \times 10^{-3}$ | 1434.5              | 1434.5            |
| Oval     | 19                | 63.8        | 154.7              | 35.5           | $2.13 \times 10^{-2}$ | 651.4               | 651.5             |



**Figure 1.** Observed (symbols) and fitted (curves) size–number cumulative relative frequency of species with: (a,b) Cardioid seeds; (c,d) elliptic seeds; (e,f) oval seeds. Left panels show the whole range of cumulative frequencies; right panels show only the two first quartiles.



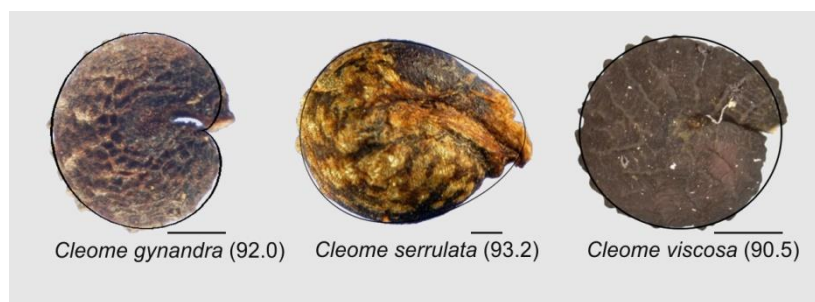
Cardioid-shaped seeds tended to be smaller and with an intermediate range of volume. Elliptic seeds tended to be larger, with the widest range of volumes and extending much further than the other two shapes at the higher end. Finally, the size of oval seeds had the lowest range extending much less than in the elliptic seeds despite surpassing them in the minimum size.

**Table 2.** Result of Scheffé’s test ( $P < 0.05$ ) showing differences in size between the seed groups.

| Model        | N  | 1     | 2     |
|--------------|----|-------|-------|
| Cardioid     | 60 | 2.4   |       |
| Oval         | 19 | 63.8  | 63.8  |
| Ellipse      | 24 |       | 139.2 |
| Significance |    | 0.431 | 0.283 |

In contrast to these differences of sizes and their ranges among shapes, values of  $c$  in the fitted Weibull equations were remarkably similar, especially those of elliptic and oval seeds,  $c = 0.491$  and  $c = 0.496$ , respectively, while cardioid seeds presented a slightly larger value of  $c = 0.535$ . Therefore, size–number distributions of seeds were highly skewed to the low end with a very pronounced tail to the right regardless of the shape, suggesting that factors governing size–number distributions of seeds, whatever they are, might be few in number or act multiplicatively or both [40].

Three *Cleome* species (*Cleome gynandra*, *Cleome serrulata* and *Cleome viscosa*) in the Cleomaceae (Brassicales) give high J index values with the three models used (cardioid, oval and ellipse; Figure 2). Of these, the seeds of *Cleome gynandra* were smaller, supporting the idea that the seeds resembling the cardioid model tended to be smaller.



**Figure 2.** *Cleome* species (*Cleome gynandra*, *Cleome serrulata* and *Cleome viscosa*) in the Cleomaceae (Brassicales) give high J index values in parentheses with cardioid, oval and ellipse models, respectively. Bars equal to 0.5 mm. *Cleome gynandra* and *Cleome serrulata* seed images are from the USDA Plants Database (<https://plants.sc.egov.usda.gov/java/>). *Cleome viscosa* seed image is from Seeds of South Australia (<https://spapps.environment.sa.gov.au/SeedsOfSA/home.html>).

#### 4. Discussion

The dataset of seed shape that was developed in this study was designed to contain those seeds resembling geometric figures. Thus, with few exceptions, it did not include seeds with wings, trichomes and other morphological adaptations for dispersal by specific mechanisms. One exception is *Nerium oleander*, whose elliptic seeds [41] have a crest of long hairs helping seed dispersal by air and water. The similarity of seeds with geometric figures allowed quantification of seed shape based on high J index values. This permitted unequivocal classification of seed types that is required for the study of the interaction between seed shape and other factors, such as seed size, plant size, life cycle and others.

The dataset contained three main morphological types based on three geometric figures: Ellipse, oval and cardioid. We prefer to use the term “cardioid” with respect to “heart-shaped” or “reniform”, both expressions are often used in botanical descriptions, because the comparison with the geometrical figure allows quantification.

The elliptic shape is common in large seeds of palms in the Arecales where it is often associated with transport of the seeds by water, but also small-sized seeds can survive a long time in water [42] and the elliptic shape associated with a smooth surface helps the intestinal transit in the cases of zoochory, like in many palms as *Phoenix dactylifera* (Arecaceae) or *Elaeagnus* (Elaeagnaceae) seeds. The combination of elliptic shape with smaller size results in seeds apt for anemochory or hydrochory like those in the Campanulaceae [43].

Most of the elliptic-shaped seeds in the dataset were intermediate-sized in the Cucurbitaceae (six species), Fagaceae (four species), Euphorbiaceae (two species) or smaller like those in the Campanulaceae (six species).

Of the 46 species analyzed having oval-shaped seeds, ten belonged to the Cucurbitaceae [23], and two (*Cucurbita maxima*, *Echinocystis lobata*) were the largest seeds in the oval-shaped species considered. Small-sized, oval-shaped seeds belonged to a diversity of species in the families Asteraceae, Cleomaceae, Cyperaceae, Rosaceae and others, illustrating the fact that for each morphological type, there is more taxonomic diversity among small- than in large-sized seeds. Also, small seeds from species of the same family, or even the same genus, may have different shapes with more frequency than large seeds. This is the case for example with the Cyperaceae, where *Cyperus iria* has small oval-shaped seeds while other species of *Carex* and *Cyperus* have ovoid-like seeds. Shape variation in taxonomic groups occurs more frequently in small seeds and may indicate a path to the genetic analysis of shape.

Cardioid shape was common in small-sized seeds, produced in high numbers by small annual or perennial herbs, such as many species in the families of the Papaveraceae in the Ranunculales [20], Malvaceae in the Malvales [21] and several families in the Caryophyllales [22], as well as the seeds of the model plant *Arabidopsis thaliana* [15] and the model legumes *Lotus japonicus* and *Medicago truncatula* [16,17]. The cardioid shape represents the growth of an organ from a fixed point [44] which in the seed is the hilum, the point of attachment of the funiculus, consistent with close dependence of the seed structure on the maternal tissue through development in small-sized seeds. Three *Cleome* species (*Cleome ginandra*, *Cleome serrulata* and *Cleome viscosa*) in the Cleomaceae (Brassicales) gave high J index values with each of the models used (cardioid, oval and ellipse, respectively). Together with species of *Lotus* and *Medicago*, two genera in which there is also seed shape variation among species [16,17], *Cleome* species may be interesting model systems for the study of the genetic determination of seed shape.

## 5. Conclusions

A dataset of seed shape is presented containing 239 species of angiosperms whose seeds resemble geometric figures (cardioid, ellipse, Fibonacci's spiral, heart shape, lens and oval). The dataset included only seeds whose images gave J index values (percent of similarity with the geometric figure) >90. The main figures were the cardioid, ellipse and oval, containing cumulatively more than 95% of the entries in the dataset. The cardioid shape was more frequent in smaller seeds while the ellipse corresponded to larger seeds. New entries will be added to the dataset as well as information including data about lifecycles, geographic distribution and other plant characters will be included to study the correlation between seed shape and lifestyle in the diverse taxonomic categories and species.

**Supplementary Materials:** The following are available online at <http://www.mdpi.com/2311-7524/5/4/71/s1>, Table S1: Seed shape database.

**Author Contributions:** Conceptualization, E.C. and L.S.D.; data curation, E.C., J.J.M.G., D.G.d.P. and L.S.D.; formal analysis, E.C., J.J.M.G., D.G.d.P. and L.S.D.; investigation, E.C., J.J.M.G., D.G.d.P. and L.S.D.; methodology, E.C., J.J.M.G., D.G.d.P. and L.S.D.; resources, E.C., J.J.M.G., D.G.d.P. and L.S.D.; software, J.J.M.G. and L.S.D.; supervision, E.C.; validation, E.C., J.J.M.G., D.G.d.P. and L.S.D.; visualization, E.C.; writing—original draft, E.C. and L.S.D.; writing—review & editing, E.C., J.J.M.G., D.G.d.P. and L.S.D.

**Funding:** This research received no external funding.

**Conflicts of Interest:** The authors declare no conflict of interest.



## References

1. Stearn, W.T. *Botanical Latin*, 2th ed.; Timber Press: Portland, OR, USA, 2004.
2. Delorit, R.J.; Gunn, C.R. *Seeds of Continental United States Legumes (Fabaceae)*; Agronomy Publications: River Falls, WI, USA, 1986.
3. Vasconcellos, J.C. *Sementes Estranhas do Trigo*; Federação Nacional dos Produtores de Trigo: Lisboa, Portugal, 1968.
4. Rodríguez, A.F.M. Morfología de las semillas de las especies del género *Trifolium* de la Península Ibérica. *Lazaroa* **1995**, *15*, 131–144.
5. Kissmann, K.G.; Groth, D. *Plantas Infestantes e Nocivas. Tomo II. Plantas Dicotiledôneas por Ordem Alfabética de Famílias: Acanthaceae a Fabaceae*, 2th ed.; BASF S.A.: São Paulo, Brazil, 1999.
6. Oval—From Wolfram MathWorld. Available online: <https://mathworld.wolfram.com/Oval.html> (accessed on 19 August 2019).
7. Ellipse—From Wolfram MathWorld. Available online: <https://mathworld.wolfram.com/Ellipse.html> (accessed on 19 August 2019).
8. Piriform Curve—From Wolfram MathWorld. Available online: <https://mathworld.wolfram.com/PiriformCurve.html> (accessed on 19 August 2019).
9. Heart Curve—FROM Wolfram MathWorld. Available online: <https://mathworld.wolfram.com/HeartCurve.html> (accessed on 19 August 2019).
10. Sun, P.; Zhang, W.; Wang, Y.; He, Q.; Shu, F.; Liu, H.; Wang, J.; Wang, J.; Yuan, L.; Deng, H. OsGRF4 controls grain shape, panicle length and seed shattering in rice. *J. Integr. Plant Biol.* **2016**, *58*, 836–847. [CrossRef] [PubMed]
11. Hu, Z.; Zhang, H.; Kan, G.; Ma, D.; Zhang, D.; Shi, G.; Hong, D.; Zhang, G.; Yu, D. Determination of the genetic architecture of seed size and shape via linkage and association analysis in soybean (*Glycine max* L. Merr.). *Genetica* **2013**, *141*, 247–254. [CrossRef]
12. Tanabata, T.; Shibaya, T.; Hori, K.; Ebana, K.; Yano, M. SmartGrain: High-Throughput Phenotyping Software for Measuring Seed Shape through Image Analysis. *Plant Physiol* **2012**, *160*, 1871–1880. [CrossRef] [PubMed]
13. Bekker, R.M.; Bakker, J.P.; Grandin, U.; Kalamees, R.; Milberg, P.; Poschlod, P.; Thompson, K.; Willems, J.H. Seed size, shape and vertical distribution in the soil: Indicators of seed longevity. *Funct. Ecol.* **1998**, *12*, 834–842. [CrossRef]
14. Cervantes, E.; Martín Gómez, J.J. Seed shape description and quantification by comparison with geometric models. *Horticulturae* **2019**, *5*, 60. [CrossRef]
15. Cervantes, E.; Martín-Gómez, J.J.; Ardanuy, R.; de Diego, J.G.; Tocino, Á. Modeling the *Arabidopsis* seed shape by a cardioid: Efficacy of the adjustment with a scale change with factor equal to the Golden Ratio and analysis of seed shape in ethylene mutants. *J. Plant Physiol.* **2010**, *67*, 408–410. [CrossRef]
16. Cervantes, E.; Martín-Gómez, J.J.; Chan, P.K.; Gresshoff, P.M.; Tocino, Á. Seed shape in model legumes: approximation by a cardioid reveals differences in ethylene insensitive mutants of *Lotus japonicus* and *Medicago truncatula*. *J. Plant Physiol.* **2012**, *169*, 1359–1365. [CrossRef]
17. Cervantes, E.; Saadaoui, E.; Tocino, Á.; Martín Gómez, J.J. Seed shape quantification in the model legumes: Methods and applications. In *The Model Legume Medicago Truncatula*; de Bruijn, F., Ed.; Wiley-Blackwell: Hoboken, NJ, USA, 2019, in press.
18. Saadaoui, E.; Martín-Gómez, J.J.; Cervantes, E. Intraspecific Variability of Seed Morphology in *Capparis spinosa* L. *Acta Biol. Cracov. Ser. Bot.* **2013**, *55*, 99–106. [CrossRef]
19. Cervantes, E.; Martín-Gómez, J.J.; Saadaoui, E. Updated Methods for Seed Shape Analysis. *Scientifica (Cairo)* **2016**, *2016*, 5691825. [CrossRef]
20. Martín-Gómez, J.J.; Rewicz, A.; Cervantes, E. Seed Shape Variability in the Order Ranunculales. *Phytotaxa* **2019**, in press.
21. Martín-Gómez, J.J.; Gutiérrez del Pozo, D.; Cervantes, E. Seed shape quantification in the Malvaceae reveals cardioid-shaped seeds predominantly in herbs. *Bot. Lith.* **2019**, *25*, 21–31. [CrossRef]
22. Martín-Gómez, J.J.; Gutiérrez del Pozo, D.; Cervantes, E. Seed shape quantification by a cardioid model in species of the order Caryophyllales: Relationship of seed shape with life style. 2019; in preparation.
23. Cervantes, E.; Martín-Gómez, J.J. Seed shape quantification in the order Cucurbitales. *Mod. Phytomorphol.* **2018**, *12*, 1–13. [CrossRef]

24. Ganhão, E.; Dias, L.S. Seed Volume Dataset—An Ongoing Inventory of Seed Size Expressed by Volume. *Data* **2019**, *4*, 61. [CrossRef]
25. Sonkoly, J.; Deák, B.; Valkó, O.; Molnár, V.A.; Tóthmérész, B.; Török, P. Do large-seeded herbs have a small range size? The seed mass-distribution range trade-off hypothesis. *Ecol. Evol.* **2017**, *7*, 11204–11212. [CrossRef] [PubMed]
26. Liu, K.; Eastwood, R.J.; Flynn, S.; Turner, R.M.; Stuppy, W.H. Seed Information Database. (Release 7.1, May 2008). About the Seed Information Database (SID) from Kew Royal Botanic Gardens. Available online: <http://data.kew.org/sid/about.html> (accessed on 23 August 2019).
27. Csontos, P.; Tamás, J.; Balogh, L. Thousand-seed weight records of species from the flora of Hungary, I. Monocotyledonopsida. *Stud. Bot. Hung.* **2003**, *34*, 121–126.
28. Csontos, P.; Tamás, J.; Balogh, L. Thousand-seed weight records of species from the flora of Hungary, II. Dicotyledonopsida. *Stud. Bot. Hung.* **2007**, *38*, 179–189.
29. Kleyer, M.; Bekker, R.M.; Knevel, I.C.; Bakker, J.P.; Thompson, K.; Sonnenschein, M.; Poschlod, P.; van Groenendael, J.M.; Klimeš, L.; Klimešová, J.; et al. The LEDA Traitbase: A database of life-history traits of Northwest European flora. *J. Ecol.* **2008**, *96*, 1266–1274. [CrossRef]
30. Török, P.; Miglécz, T.; Valkó, O.; Tóth, K.; Kelemen, A.; Albert, Á.-J.; Matus, G.; Molnár, V.A.; Ruprecht, E.; Papp, L.; et al. New thousand-seed weight records of the Pannonian flora and their application in analysing social behaviour types. *Acta Bot. Hung.* **2013**, *55*, 429–472. [CrossRef]
31. Török, P.; Tóth, E.; Tóth, K.; Valkó, O.; Deák, B.; Kelbert, B.; Bálint, P.; Radóczy, S.Z.; Kelemen, A.; Sonkoly, J.; et al. New measurements of thousand-seed weights of species in the Pannonian flora. *Acta Bot. Hung.* **2016**, *58*, 187–198. [CrossRef]
32. Moles, A.T.; Ackerly, D.D.; Tweddle, J.C.; Dickie, J.B.; Smith, R.; Leishman, M.R.; Mayfield, M.M.; Pitman, A.; Wood, J.T.; Westoby, M. Global patterns in seed size. *Glob. Ecol. Biogeogr.* **2007**, *16*, 109–116. [CrossRef]
33. World Flora Online. Available online: <https://www.worldfloraonline.org/> (accessed on 19 May 2019).
34. Weibull, W. A statistical distribution function of wide applicability. *J. Appl. Mech.* **1951**, *18*, 293–297.
35. Dubey, S.D. Normal and Weibull distributions. *Nav. Res. Logist. Q.* **1967**, *14*, 69–79. [CrossRef]
36. Bonner, F.T.; Dell, T.R. The Weibull function: A new method of comparing seed vigor. *J. Seed. Technol.* **1976**, *1*, 96–103.
37. Marquardt, D.W. An algorithm for least-squares estimation of nonlinear parameters. *J. Soc. Ind. Appl. Math.* **1963**, *11*, 431–441. [CrossRef]
38. Martín-Gómez, J.J.; Rewicz, A.; Goriewa-Duba, K.; Wiwart, M.; Tocino, Á.; Cervantes, E. Morphological Description and Classification of Wheat Kernels Based on Geometric Models. *Agronomy* **2019**, *9*, 399. [CrossRef]
39. Thompson, K.; Band, S.R.; Hodgson, J.G. Seed shape and size predict persistence in soil. *Funct. Ecol.* **1993**, *7*, 236–241. [CrossRef]
40. Limpert, E.; Stahel, W.A.; Abbt, M. Log-normal distributions across the sciences: Keys and clues. *BioScience* **2001**, *51*, 341–352. [CrossRef]
41. Saadaoui, E.; Martín-Gómez, J.J.; Ben Yahia, K.; Cervantes, E. Seed Diversity and Germination in Spontaneous and Cultivated Populations of *Nerium oleander* var. Villa Romaine Grown in Tunisia. *Not. Bot. Hort. Agrob.* **2019**, *47*, in press.
42. Wang, W.; Chen, Q.; Chen, K. Seed bank characteristics of the *Nymphoides peltata* population in Lake Taihu. *Sci. Rep.* **2015**, *5*, 13261. [CrossRef]
43. Maier, A.; Emig, W.; Leins, P. Dispersal Patterns of some *Phyteuma* Species (Campanulaceae). *Plant Biol.* **1999**, *1*, 408–417. [CrossRef]
44. Thompson, D. *On Growth and Form*; Cambridge University Press: London, UK, 1917; 1116p.

