Posted on 20 Mar 2023 | The copyright holder is the author/funder. All rights reserved. No reuse without permission. | https://doi.org/10.22541/au.167935473.32803184/v1 | This a preprint and has not been peer reviewed. Data may be preliminary.

DiasMorph: a dataset of morphological traits and images of Central European diaspores

Roberta L. C. Dayrell[1], Tankred Ott[1], Lina Begemann[1], and Peter Poschlod[1]

[1]Affiliation not available

March 20, 2023

Abstract

A knowledge of the morphology of seeds and diaspores is an invaluable resource that has prompted the compilation of descriptions and/or images in a variety of media, including databases with standardized protocols for trait measurement. The pressing need for new solutions to solve environmental crises, together with advances in the application of machine learning and image analysis to solve ecological problems, call for an upgrade of the diaspore morphological datasets. Here, we present DiasMorph, a dataset of morphological traits and images of diaspores from 1,437 taxa in 513 genera, and 96 families from Central Europe, totalling 94,214 records. Moreover, DiasMorph compiles information on quantitative morphological traits extracted from the images following a standardized and reproducible image analysis method, and include not only traditional morphometric measurements, but also colour, and shape features made available for the first time in a large dataset. The quantitative traits records correspond to measurements of individual diaspores, an input currently unavailable in traits databases that will allow for several approaches to be used for a complete exploration of the morphological traits of these species. We also included information on the presence and absence of appendages and structures in the diaspores of the evaluated taxa. By making these data available, we aim to encourage initiatives to advance on new tools for diaspore identification, further our understanding of morphological traits functions, and provide means for the continuous development of image analyses applications.

Keywords: diaspore appendages, dispersal unit, image dataset, plant traits dataset, seed colour, seed morphology, seed shape, seed size, seed traits.

DiasMorph: a dataset of morphological traits and images of Central European diaspores

Roberta L. C. Dayrell[1,2]*, Lina Begemann[1], Tankred Ott[1], Peter Poschlod[1]

[1] Faculty of Biology and Preclinical Medicine, University of Regensburg, Universitätsstraße 31, Regensburg, D-93053, Germany.

[2] Royal Botanic Gardens, Kew, Wakehurst, Ardingly, Haywards Heath, West Sussex, RH17 6TN, UK.

* Corresponding author: Roberta L. C. Dayrell (dayrell.rlc@gmail.com)

Open Research Statement: Data will be made available on Figshare repository upon acceptance of the method paper (this paper has been temporarily referred to as Dayrell et al. *submitted*).

Introduction

The morphological description of seeds and diaspores offers essential information for scientists and practitioners in a wide variety of fields, including botany, restoration, conservation, ethnobotany, archaeology, and agriculture. Diaspore traits, such as size, shape, colour, surface structures, and the presence of appendages are needed to establish the identity of particular diaspores that become detached of their mother plant (Martin and Barkley 1961), for instance in seed lots, seed traps, soil seed bank, archaeological sites, or forensic investigations. Moreover, integrating diaspore morphological traits into theoretical plant regeneration

framework can lead to major advances in predictive evolutionary and ecological models, and thereby support conservation and restoration actions (Saatkamp et al. 2019).

Throughout the years, the demand for knowledge of diaspore morphology has led to numerous compilations of text descriptions and/or images of diaspores in books, guides and atlases (e.g., Martin and Barkley 1961, Brouwer and Stählin 1975, Beijerinck 1976, Sweedman and Merritt 2006, Bojňanský and Fargašová 2007, Cappers et al. 2012). In the last two decades, databases have been built to synthesise and centralise information on diaspore traits (e.g., Kleyer et al. 2008, Hintze et al. 2013, GEVES 2022, Royal Botanic Gardens Kew 2022), facilitating large scale analyses. Along with databases, standardized protocols were established for trait measurements to allow for the integration of data with different sources. These included methods for the description of diaspores, which consist of the quantification of size and other morphometric measurements (most reported as taxa mean or range values), and the classification of attributes either based on visual (perceptual) categories and/or functional structures and/or anatomical parts (Römermann et al. 2005).

Recently, the pressing need for new solutions to deal with environmental crises, together with the surge in applications of machine learning and image analysis in ecology and evolution, calls for an upgrade of the diaspore morphological datasets. The automated extraction of information from digital images provides the opportunity to collect quantitative phenotypic data in large quantities, enabling the investigation of high dimensional and complex relationships between traits and their interaction with environmental variables (Lürig et al. 2021). Furthermore, the use of machine learning algorithms to classify images and/or suites of traits can allow for the automation of taxa identification, making the task faster and not exclusively dependent on experienced taxonomists (Borowiec et al. 2022, Loddo et al. 2022).

Here, we present DiasMorph, a comprehensive dataset of morphological traits and images of diaspores from Central Europe. It provides images of 94,214 diaspores from 1,437 taxa in 513 genera, and 96 families, captured with a standardized and reproducible method (Dayrell et al.*submitted*). The dataset also compiles information on quantitative morphological traits extracted from the images following an image analysis method and include not only traditional morphometric measurements, but also colour, and shape features made available for the first time in a large dataset (Dayrell et al. *submitted*). The quantitative traits records correspond to measurements of individual diaspores, an input currently unavailable in trait databases that will allow for several approaches to be used for a complete exploration of the morphological traits of these species. We also included information on the presence and absence of appendages and structures in the diaspores of the evaluated taxa. By making these data available, we aim to encourage initiatives to advance on new tools for diaspore identification, further our understanding of morphological traits functions, and provide means for the continuous development of image analyses applications.

Metadata

Class I. Data set descriptors

A. Data set identity

Title: DiasMorph: a dataset of morphological traits and images of Central European diaspores

B. Data set identification code

Suggested Data set Identity Code: DiasMorph

C. Data set description

The present dataset contains images and records of quantitative morphological traits for 94,214 diaspores from 1,437 taxa (including species, lower ranks, and three sections) in 513 genera, and 96 families from Central Europe kept in the seed collection of the Chair of Ecology and Conservation Biology at the University of Regensburg, Germany. The most represented families in the database are Asteraceae (191 taxa; 65 genera), Poaceae (113; 48), Brassicaceae (93; 43), Cyperaceae (85; 11), and Fabaceae (80; 22), with ~39% of all reported species (Figure 1). Images from diaspores were obtained with a flatbed scanner in a high-contrast

Posted on 20 Mar 2023 | The copyright holder is the author/funder. All rights reserved. No reuse without permission. | https://doi.org/10.22541/au.167935473.32803184/v1 | This a preprint and has not been peer reviewed. Data may be preliminary.

Posted on 20 Mar 2023 | The copyright holder is the author/funder. All rights reserved. No reuse without permission. | https://doi.org/10.22541/au.167935473.32803184/v1 | This a preprint and has not been peer reviewed. Data may be preliminary.

background and pre-processed for colour standardization; quantitative morphological traits were extracted from the images (Dayrell et al. *submitted*). The morphological traits include: 1) morphometric measurements (length, width, area, perimeter, aspect ratio, circularity, surface structure, and solidity); 2) colour measurements for human recognition purposes (Figure 2; mean, median, and most dominant colours in sRGB), and ecological and evolutionary studies (independent of any particular animal visual system; linear sRGB); 3) standardized contour of diaspores (50 coordinates for each seed) for shape analysis methods. Diaspores of the evaluated taxa were classified according to the presence and absence of appendages and structures (e.g., fleshy structures, elongated appendages, surface hairs).

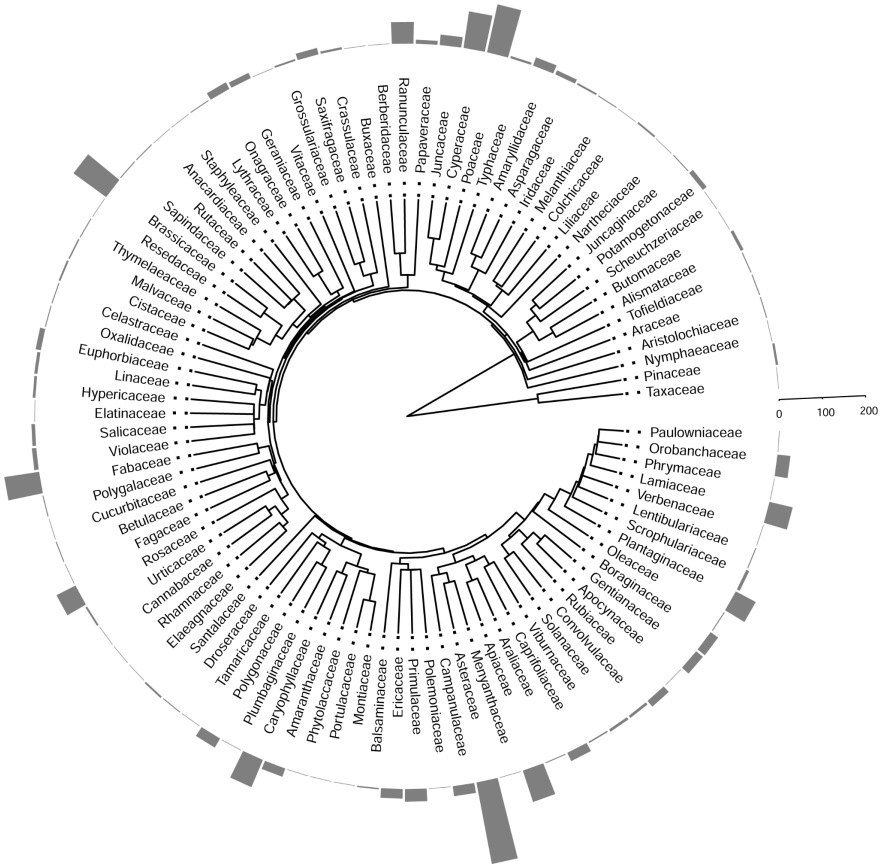


Figure 1. Cladogram of the phylogeny for the families in DiasMorph dataset. The barplot represents the number of taxa within each family in the DiasMorph dataset.

Posted on 20 Mar 2023 | The copyright holder is the author/funder. All rights reserved. No reuse without permission. | https://doi.org/10.22541/au.167935473.32803184/v1 | This a preprint and has not been peer reviewed. Data may be preliminary.

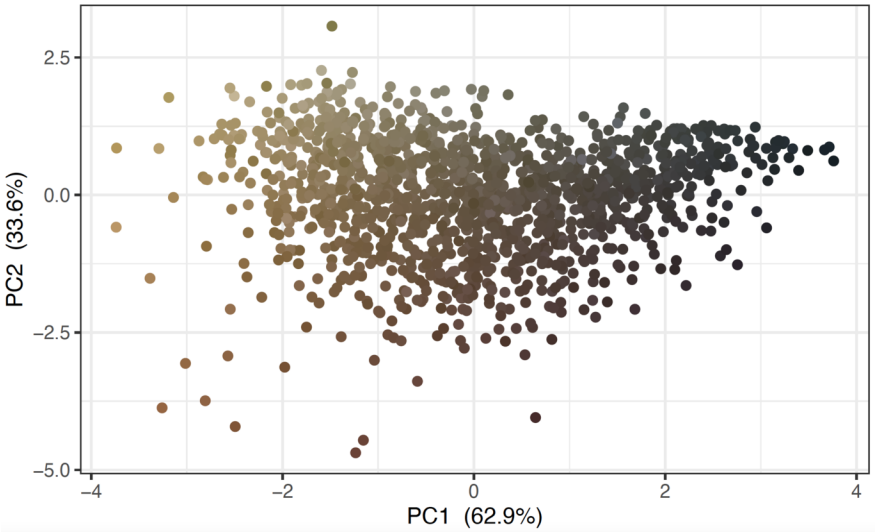


Figure 2. PCA scores plot obtained from the median colour values of diaspores from the 1,437 taxa in the DiasMorph dataset.

Principal investigators:

Roberta L. C. Dayrell. Faculty of Biology and Preclinical Medicine, University of Regensburg, Universitätsstraße 31, Regensburg, D-93053, Germany.

Peter Poschlod. Faculty of Biology and Preclinical Medicine, University of Regensburg, Universitätsstraße 31, Regensburg, D-93053, Germany.

D. Keywords: diaspore appendages, dispersal unit, image dataset, plant traits dataset, seed colour, seed morphology, seed shape, seed size, seed traits.

Class II. Research origin descriptors

A. Overall project description

1. Identity: Morphological traits and images of diaspores from Central Europe.

2. Originators: The project "DiasMorph: a dataset of morphological traits and images of Central European diaspores" was carried out at the University of Regensburg, Germany as part of the project "Archeoplant: research on plant identification using artificial intelligence".

3. Period of study: NA

4. Objectives: To build a comprehensive dataset of images and morphological traits of Central European diaspores that can be used to advance on new tools for diaspore identification, further our understanding of morphological traits functions, and provide means for the continuous development of image analyses applications.

5. Abstract: same as above.

6. Sources of funding: This research is supported by the European Regional Development Fund (ERDF) and by the Target ETC Program Free State of Bavaria – Czech Republic 2014-2020 (INTERREG V, Project 343).

B. Specific subproject description

Posted on 20 Mar 2023 | The copyright holder is the author/funder. All rights reserved. No reuse without permission. | https://doi.org/10.22541/au.167935473.32803184/v1 | This a preprint and has not been peer reviewed. Data may be preliminary.

1. Network establishment: The DiasMorph network was established in September of 2021, led by Roberta L. C. Dayrell and Peter Poschlod.

2. Site description: NA.

3. Data sampling:

3.1 Sampled taxa

We sampled diaspores of Central European species kept in the seed collection of the Chair of Ecology and Conservation Biology at the University of Regensburg, Germany, started and curated by Prof. Peter Poschlod. In total, our dataset included 1,437 taxa (including species, lower ranks, and three sections), belonging to 513 genera, 96 plant families, making up a total of 94,214 diaspores. (*See* Taxonomic data section for information on nomenclature verification). The most represented families in the database are Asteraceae (191 taxa; 65 genera), Poaceae (113; 48), Brassicaceae (93; 43), Cyperaceae (85; 11), and Fabaceae (80; 22), with ~39% of all reported species.

3.2 Recorded appendages

For each species, we recorded diaspore structures and appendages (Table 1) following a modified version of seed structure categories in LEDA Trait standards (Römermann et al. 2005, Kleyer et al. 2008). As the LEDA is a database focused on functional traits, the modifications aimed to improve the objectivity of the classification and facilitate the recognition of morphological structures for identification purposes. For each taxon, appendages and structures were classified as present (1) or absent (0). In some instances, diaspores of species and genera were found with and without appendages and structures; for these cases, we recorded the structures as present and later specified them as missing from the image (*see* Section 3.3).

Table 1. Summary of the diaspore appendage and structure categories.

Appendage/Structure	Description
1. Fleshy cover	Fleshy structure (e.g., pulp), or cup-like structure (e.g., aril) that totally or partly covers the s
2. Fleshy appendage	Fleshy appendage or structure (e.g., elaiosome) attached to seed. It can vary in form, size, and
3. Dry covering structure	Dry covering structures that partially or completely cover the germination unit. Presence was
4. flat appendage	Membranous / delicate structures that stick out of the more compact part of the germination u
5. hairy appendage	A tuft, hairlike branches, or ring of hairs (and less often scales) which are attached to the germ
Elongated appendages	Structures that prominently stick out of the main part of the diaspore, having a length conside
6.1 Elongated short	Elongated appendage is at least one tenth of the length and shorter than half of the diaspore's
6.2 Elongated long	Elongated appendage is at least half of the length of the diaspore.
6.3 Elongated spiral coiled	Spiral coiled elongated appendage.
6.4 Elongated bent	Elongated appendage is bent or distinctively crooked.
6.5 Elongated hairy	Elongated appendage has fine, flexible, linear outgrowths. Different from hairy appendage as t
6.6 Elongated bristles	Elongated appendage has linear outgrowths that are semiflexible, thicker and generally shorter
6.7 Elongated multiple	Two or more elongated appendages are present.
7.1 single hook	One elongated appendage with very recurved beak in a hook-like shape.
7.2 multiple hooks	Bristles or spines with curved or backwards pointing tips, or with secondary bristles along thei
8.1 surface hairs	Fine, flexible, linear outgrowths on the surface.
8.2 surface bristles	Linear outgrowths on the surface that are semiflexible, thicker and generally shorter than hairs

3.3 Extraction of quantitative traits

We used an image analysis method described in detail by Dayrell et al. (*submitted*) to obtain images and extract quantitative measurements of diaspore morphology.

Sample preparation

We cleaned the diaspores with the aid of a stereo microscope and only selected diaspores with all structures in a well-preserved state, apart from three exceptions. 1) Fleshy covering structures and some fleshy outgrowths were removed due to the pronounced changes that these structures undergo after dispersal, which can lead to unrecognisable colours, shapes, and sizes. 2) We measured diaspores without scales or covering structures when most diaspores in a vial of the seed collection had detached from these structures without handling. 3) Hairy appendages (e.g., pappus and plumes) were removed due to requirements of the method (details in Dayrell et al. *submitted*). The structures that were not present in the scanned diaspores were recorded as 'missing structures' in the dataset.

Image acquisition

For image acquisition, diaspores were arranged on the flat scanner avoiding any contact or overlap. The number of sampled diaspores varied for each taxon according to their availability in the seed collection (Figure 3). We sampled all available material that met sample preparation standards for up to 30 diaspores. In cases where the number of available diaspores exceeded 30, we sampled seeds to cover an area of up to 200 cm^2. The flatbed scanner was covered with a wooden frame 10 mm thick with a royal blue background (Figure 4). Images were acquired with a flatbed scanner (HP Scanjet G4010) at a resolution of 1,200 DPI to well-represent small seeds and fine appendages. All automatic correction functions associated with the scanner software were disabled to ensure that the RGB values of the samples were not manipulated. The resulting images were saved in the Joint Photographic Experts Group (JPEG) format with no compression.

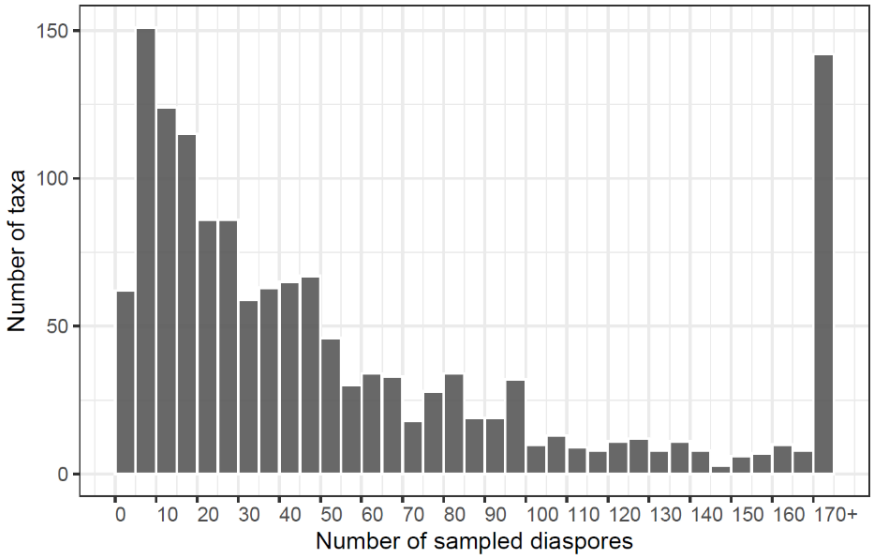


Figure 3. Histogram of the number of diaspores per taxon sampled for quantitative measurements.

Posted on 20 Mar 2023 | The copyright holder is the author/funder. All rights reserved. No reuse without permission. | https://doi.org/10.22541/au.167935473.32803184/v1 | This a preprint and has not been peer reviewed. Data may be preliminary.

Posted on 20 Mar 2023 | The copyright holder is the author/funder. All rights reserved. No reuse without permission. | https://doi.org/10.22541/au.167935473.32803184/v1 | This a preprint and has not been peer reviewed. Data may be preliminary.

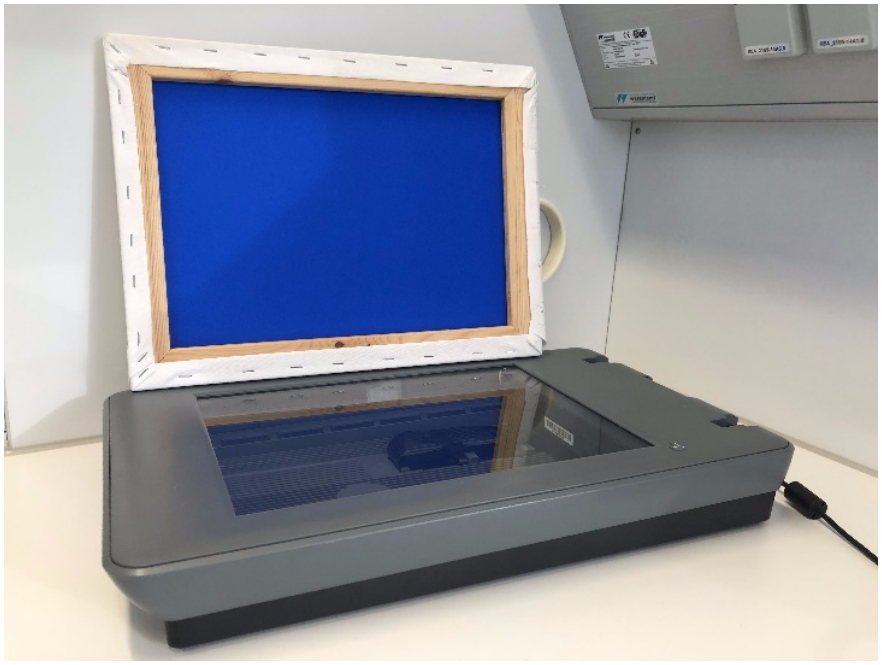


Figure 4. The flatbed scanner and a wooden frame 10 mm thick covered with a royal blue cardboard, which was used to provide the high-contrast background.

Image processing

To allow standardisation of colour measurements, a Spyder Checkr[?] 24 card (Datacolor, NJ, USA) was scanned in the flatbed scanner under the same settings as the diaspores, and used to calculate a colour conversion matrix (CCM). The CCM was then applied to images for optimal colour reproduction (https://github.com/rdayrell/colour_calibration). In some images, undesired elements, such as broken seeds and particles, were removed from image with the brush and clone stamp tools in Adobe Photoshop. Images were saved in PNG format throughout all processing steps to avoid compression artifacts. Processed images (Figure 5) comprise the image dataset and were used as inputs for automated trait extraction.

Figure 5. Examples of diaspore images in the DiasMorph dataset.

Extraction with Traitor software

The Traitor software https://github.com/TankredO/traitor was used to segment, align, and extract morphological traits from images (Dayrell et al. *submitted*). The extracted traits include morphometric measurements (length, width, aspect ratio, area, perimeter, diaspore surface structure, solidity, circularity"), colour measurements (sRGB, linear sRGB, dominant colours), and 50 coordinates of the shape outline standardised for size invariance.

Data quality assessment

The only systematic error found by the manual inspection was an alignment failure in diaspores with sharply bent awns belonging from the Poaceae family, which resulted in inaccurate length and width measurements. To correct for this error, we edited images of Poaceae species and erased the elongated appendage with the brush and clone stamp tools in Adobe Photoshop, saved it as PNG. The images were labelled with the same

name as the original image, with the addition of '_edit' (e.g., 'img_0261' and 'img_0261_edit'). Traits of edited images were extracted with Traitor and merged with the dataset with the addition of 'elongated appendages' on the 'missing structures' column. Original records of species with bent appendages were deleted from the dataset, while those of species with unbent appendages were kept.

4. Taxonomic data: Species names and family information were using the functions *WFO.match and WFO.one* from the R package *WorldFlora* (Kindt 2020). The last nomenclature verification was carried out on September 16th, 2022.

Class III. Data set status and accessibility

A. Status

Latest update: December 14th, 2022.

Latest archive date: Data and metadata will be made available on Figshare upon acceptance of Dayrell et al. (*submitted*).

Metadata status: Last update, March 14th, 2023.

B. Accessibility:

Storage location: Figshare repository - https://doi.org/10.6084/m9.figshare.21206507.v2

Contact person: Roberta L. C. Dayrell. E-mail: dayrell.rlc@gmail.com

Copyright restrictions: None.

Proprietary restrictions: Please cite this data paper when using the current data in publications and let us know how the data is used in the publications.

Costs: None.

Class IV. Data structural descriptors

A. Data set file

Identity and size:

- DiasMorph_original_images.zip (1,547 zipped image files), 10.48 GB.
- DiasMorph_edited_images.zip (41 zipped image files), 226.4 MB.
- sample_images.zip (4 sample image files), 36.57 MB.
- DiasMorph_labels_and_structures.csv, 197 KB.
- DiasMorph_quantitative_traits.csv, 214.2 MB.

Format and storage mode: All image (.png) are included in compressed folders (.zip). Data tables are comma-separated values (.csv) files.

Header information: See column descriptions in Metadata worksheet.

Alphanumeric attributes: Mixed

Data anomalies: Missing information was classified as "NA".

B. Variable information

Description for files in DiasMorph_original_images.zip.

The zip file contains 1,547 colour images of 1,437 taxa in PNG format and 1,200 DPI resolution. Images were obtained with a flatbed scanner and subjected to colour standardisation (*see* Section 3).

Description for files in DiasMorph_edited_images.zip.

Posted on 20 Mar 2023 | The copyright holder is the author/funder. All rights reserved. No reuse without permission. | https://doi.org/10.22541/au.167935473.32803184/v1 | This a preprint and has not been peer reviewed. Data may be preliminary.

Posted on 20 Mar 2023 | The copyright holder is the author/funder. All rights reserved. No reuse without permission. | https://doi.org/10.22541/au.167935473.32803184/v1 | This a preprint and has not been peer reviewed. Data may be preliminary.

The zip file contains 41 colour images of 34 taxa in PNG format and 1,200 DPI resolution. Original images were edited as described under 'Data quality assessment' subheading (*see* Section 3).

Description for files in DiasMorph_edited_images.zip**.**

The zip file contains four sample images included in 'DiasMorph_original_images.zip'.

Description for dataset DiasMorph_labels_and_structures.csv.

The dataset contains image labels and the recorded diaspore structures and appendages (Table 2).

Table 2. Variable names and descriptions for file DiasMorph_labels_and_structures.csv. Definition of categories are in Table 1.

Variables	**Description**
image_name	Image file name in image dataset
scientificName	Accepted species name.
scientificNameAuthorship	Author of the species accepted name.
spec.name.ORIG	Species scientific name in the data source.
genus	Taxonomic genus.
family	Taxonomic Family.
missing_structures	Structures that are not present in the image.
c1_fleshy_cover	Absence (0), presence (1).
c2_fleshy_appendage	Absence (0), presence (1).
c3_dry_covering_structure	Absence (0), presence (1).
c4_flat_appendage	Absence (0), presence (1).
c5_hairy_appendage	Absence (0), presence (1).
c6_1_elongated_short	Absence (0), presence (1).
c6_2_elongated_long	Absence (0), presence (1).
c6_3_elongated_spiral_coiled	Absence (0), presence (1).
c6_4_elongated_bent	Absence (0), presence (1).
c6_5_elongated_hairy	Absence (0), presence (1).
c6_6_elongated_bristles	Absence (0), presence (1).
c6_7_elongated_multiple	Absence (0), presence (1).
c7_1_single_hook	Absence (0), presence (1).
c7_2_multiple_hooks	Absence (0), presence (1).
c8_1_surface_hairs	Absence (0), presence (1).
c8_2_surface_bristles	Absence (0), presence (1).
img_dataset	'original' for images in 'DiasMorph_original_images.zip' and 'edited' for images in 'DiasM

Description for dataset DiasMorph_quantitative_traits.csv.

The dataset contains image labels and the quantitative traits extracted from images.

Table 3. Variable names and descriptions for file DiasMorph_quantitative_traits.csv.

Variables	**Description**
sample_name	Name of the single diaspore image extracted by Traitor. Unique for each diaspore.
image_name	Image file name in image dataset.
scientificName	Accepted species name.
scientificNameAuthorship	Author of the species accepted name.
spec.name.ORIG	Species scientific name in the data source.

Variables	Description
genus	Taxonomic genus.
family	Taxonomic Family.
missing_structures	Structures that are not present in the diaspore image.
length	the longest dimension of the diaspore in millimetres (mm)
width	the widest axis perpendicular to the length axis in millimetres (mm)
aspect_ratio	Proportional relationship between width and length. Dimensionless. Formula: aspect ratio = width ÷ length.
area	Area of diaspore contour in square millimetres (mm^2).
perimeter	Perimeter of diaspore contour in millimetres (mm).
surface_structure	Shape descriptor that describes the roughness of a diaspore's surface. The maximum value is 1 (completely smooth), and surface is increasingly rougher as the value approaches 0. Dimensionless. Formula: surface structure = perimeter (convex hull) ÷ perimeter.
solidity	Shape descriptor that describes the overall concavity of a diaspore. The maximum value is 1 (completely convex shape), and lower values indicate greater extent of concavity and/or irregular boundaries. Dimensionless. Formula: solidity = area (convex hull) ÷ area.
circularity	Shape descriptor that describes the degree of similarity to a perfect circle. The maximum value is 1 (perfect circle), and the shape is increasingly less circular as the value approaches 0. Formula: circularity = $4\pi \times$ area / perimeter^2.
R_median	Linearized median pixel value for the red (R) channel within the sRGB colour space. Range 0 (absence of channel) to 255 (maximum value).
G_median	Linearized median pixel value for the green (G) channel within the sRGB colour space. Range 0 (absence of channel) to 255 (maximum value).
B_median	Linearized median pixel value for the blue (B) channel within the sRGB colour space. Range 0 (absence of channel) to 255 (maximum value).
R_mean	Linearized mean pixel value for the red (R) channel within the sRGB colour space. Range 0 (absence of channel) to 255 (maximum value).
G_mean	Linearized mean pixel value for the green (G) channel within the sRGB colour space. Range 0 (absence of channel) to 255 (maximum value).
B_mean	Linearized mean pixel value for the blue (B) channel within the sRGB colour space. Range 0 (absence of channel) to 255 (maximum value).

Posted on 20 Mar 2023 | The copyright holder is the author/funder. All rights reserved. No reuse without permission. | https://doi.org/10.22541/au.167935473.32803184/v1 | This a preprint and has not been peer reviewed. Data may be preliminary.

Variables	Description
sR_median	Median pixel value for the red (R) channel. sRGB colour space: range 0 (absence of channel) to 255 (maximum value).
sG_median	Median pixel value for the green (G) channel. sRGB colour space: range 0 (absence of channel) to 255 (maximum value).
sB_median	Median pixel value for the blue (B) channel. sRGB colour space: range 0 (absence of channel) to 255 (maximum value).
sR_mean	Mean pixel value for the red (R) channel. sRGB colour space: range 0 (absence of channel) to 255 (maximum value).
sG_mean	Mean pixel value for the green (G) channel. sRGB colour space: range 0 (absence of channel) to 255 (maximum value).
sB_mean	Mean pixel value for the blue (B) channel. sRGB colour space: range 0 (absence of channel) to 255 (maximum value).
rgb_0_r	Median pixel value for the red (R) channel of the second most dominant colour. sRGB colour space: range 0 (absence of channel) to 255 (maximum value).
rgb_0_g	Median pixel value for the green (G) channel of the second most dominant colour. sRGB colour space: range 0 (absence of channel) to 255 (maximum value).
rgb_0_b	Median pixel value for the blue (B) channel of the second most dominant colour. sRGB colour space: range 0 (absence of channel) to 255 (maximum value).
rgb_1_r	Median pixel value for the red (R) channel of the most dominant colour. sRGB colour space: range 0 (absence of channel) to 255 (maximum value).
rgb_1_g	Median pixel value for the green (G) channel of the most dominant colour. sRGB colour space: range 0 (absence of channel) to 255 (maximum value).
rgb_1_b	Median pixel value for the blue (B) channel of the most dominant colour. sRGB colour space: range 0 (absence of channel) to 255 (maximum value).
rgb_0_frac	Fraction of pixels assigned as the second most dominant colour. Range 0 to 0.5.
rgb_1_frac	Fraction of pixels assigned as the most dominant colour. Range 0.5 to 1.
x_0 – x_49	x values for each of the 50 xy coordinates that comprise the diaspore contour after standardization for size invariance. Columns x_0 to x_49.

Posted on 20 Mar 2023 | The copyright holder is the author/funder. All rights reserved. No reuse without permission. | https://doi.org/10.22541/au.167935473.32803184/v1 | This a preprint and has not been peer reviewed. Data may be preliminary.

Variables	Description
y_0 – y_49	y values for each of the 50 xy coordinates that comprise the diaspore contour after standardization for size invariance. Columns y_0 to y_49.

Class V. Supplemental descriptors

A. Data acquisition

1. Data request history: None.

2. Data set updates history: None.

3. Data entry/verification procedures: done by first author.

B. History of data usage: None.

Acknowledgments

We thank all colleagues and friends who contributed to the seed collection, and Mirna Bosnjakovic for help with diaspore scanning. This research is supported by the European Regional Development Fund (ERDF) and by the Target ETC Program Free State of Bavaria – Czech Republic 2014-2020 (INTERREG V, Project 343).

Literature cited

Beijerinck, W. 1976. Zadenatlas der Nederlandsche Flora. Backhuys & Meesters, Amsterdam.

Bojňanský, V., and A. Fargašová. 2007. Atlas of seeds and fruits of central and east-European flora. Springer Netherlands, Dordrecht.

Borowiec, M. L., R. B. Dikow, P. B. Frandsen, A. McKeeken, G. Valentini, and A. E. White. 2022. Deep learning as a tool for ecology and evolution. Methods in Ecology and Evolution 13:1640–1660.

Brouwer, W., and A. Stählin. 1975. Handbuch der Samenkunde für Landwirtschaft, Gartenbau und Forstwirtschaft. DLG-Verlag, Frankfurt.

Cappers, R. T., R. M. Bekker, and J. E. Jans. 2012. Digitale zadenatlas von Nederland. 2nd edition. Barkhuis Publishing & Groningen University Library, Groningen.

GEVES. 2022. I.D.SEED®. https://mediatheque.geves.fr/geves/identification.

Hintze, C., F. Heydel, C. Hoppe, S. Cunze, A. König, and O. Tackenberg. 2013. D3: The Dispersal and Diaspore Database - Baseline data and statistics on seed dispersal. Perspectives in Plant Ecology, Evolution and Systematics 15:180–192.

Kindt, R. 2020. WorldFlora: An R package for exact and fuzzy matching of plant names against the World Flora Online taxonomic backbone data. Applications in Plant Sciences 8:e11388.

Kleyer, M., R. M. Bekker, I. C. Knevel, J. P. Bakker, K. Thompson, M. Sonnenschein, P. Poschlod, J. M. van Groenendael, L. Klimeš, J. Klimešová, S. Klotz, G. M. Rusch, M. Hermy, D. Adriaens, G. Boedeltje, B. Bossuyt, A. Dannemann, P. Endels, L. Götzenberger, J. G. Hodgson, A.-K. Jackel, I. Kühn, D. Kunzmann, W. A. Ozinga, C. Römermann, M. Stadler, J. Schlegelmilch, H. J. Steendam, O. Tackenberg, B. Wilmann, J. H. C. Cornelissen, O. Eriksson, E. Garnier, and B. Peco. 2008. The LEDA Traitbase: a database of life-history traits of the Northwest European flora. Journal of Ecology 96:1266–1274.

Loddo, A., C. Di Ruberto, A. M. P. G. Vale, M. Ucchesu, J. M. Soares, and G. Bacchetta. 2022. An effective and friendly tool for seed image analysis. Visual Computer.

Posted on 20 Mar 2023 | The copyright holder is the author/funder. All rights reserved. No reuse without permission. | https://doi.org/10.22541/au.167935473.32803184/v1 | This a preprint and has not been peer reviewed. Data may be preliminary.

Lürig, M. D., S. Donoughe, E. I. Svensson, A. Porto, and M. Tsuboi. 2021. Computer Vision, Machine Learning, and the Promise of Phenomics in Ecology and Evolutionary Biology. Frontiers in Ecology and Evolution 9.

Martin, A. C., and W. D. Barkley. 1961. Seed Identification Manual. University of California Press, Berkeley and Los Angeles.

Römermann, C., O. Tackenberg, L. Götzenberg, and P. Poschlod. 2005. Morphology dispersal unit. Pages 112–118 *in* I. C. Knevel, R. M. Bekker, D. Kunzmann, M. Stadler, and K. Thompson, editors. The LEDA Traitbase Collecting and Measuring Standards. Groningen University, Groningen.

Royal Botanic Gardens Kew. 2022. Seed Information Database (SID). Version 7.1. http://data.kew.org/sid/.

Saatkamp, A., A. Cochrane, L. Commander, L. K. Guja, B. Jimenez-Alfaro, J. Larson, A. Nicotra, P. Poschlod, F. A. O. Silveira, A. T. Cross, E. L. Dalziell, J. Dickie, T. E. Erickson, A. Fidelis, A. Fuchs, P. J. Golos, M. Hope, W. Lewandrowski, D. J. Merritt, B. P. Miller, R. G. Miller, C. A. Offord, M. K. J. Ooi, A. Satyanti, K. D. Sommerville, R. Tangney, S. Tomlinson, S. Turner, and J. L. Walck. 2019. A research agenda for seed-trait functional ecology. New Phytologist 221:1764–1775.

Sweedman, L., and D. Merritt. 2006. Australian seeds: a guide to their collection, identification and biology. Page (L. Sweedman and D. Merritt, Eds.). CSIRO Publishing, Melbourne.

Posted on 20 Mar 2023 | The copyright holder is the author/funder. All rights reserved. No reuse without permission. | https://doi.org/10.22541/au.167935473.32803184/v1 | This a preprint and has not been peer reviewed. Data may be preliminary.