

Automated extraction of seed morphological traits from images

Roberta L. C. Dayrell^{1,2}  | Tankred Ott¹  | Tom Horrocks³  | Peter Poschod¹ 

¹Faculty of Biology and Preclinical Medicine, University of Regensburg, Regensburg, Germany

²Royal Botanic Gardens, Kew, Wakehurst, Ardingly, Haywards Heath, West Sussex, UK

³School of Earth Sciences, University of Western Australia, Crawley, Western Australia, Australia

Correspondence

Roberta L. C. Dayrell
Email: dayrell.rlc@gmail.com

Funding information

European Regional Development Fund, Grant/Award Number: Project 343

Handling Editor: Arthur Porto

Abstract

1. The description of biological objects, such as seeds, mainly relies on manual measurements of few characteristics, and on visual classification of structures, both of which can be subjective, error prone and time-consuming. Image analysis tools offer means to address these shortcomings, but we currently lack a method capable of automatically handling seeds from different taxa with varying morphological attributes and obtaining interpretable results.
2. Here, we provide a simple image acquisition and processing protocol and introduce *Traitor*, an open-source software available as a command-line interface (CLI), which automates the extraction of seed morphological traits from images. The workflow for trait extraction consists of scanning seeds against a high-contrast background, correcting image colours, and analysing images with the software. *Traitor* is capable of processing hundreds of images of varied taxa simultaneously with just three commands, and without a need for training, manual fine-tuning or thresholding. The software automatically detects each object in the image and extracts size measurements, traditional morphometric descriptors widely used by scientists and practitioners, standardised shape coordinates, and colorimetric measurements.
3. The method was tested on a dataset comprising of 91,667 images of seeds from 1228 taxa. *Traitor's* extracted average length and width values closely matched the average manual measurements obtained from the same collection (concordance correlation coefficient of 0.98). Further, we used a large image dataset to demonstrate how *Traitor's* output can be used to obtain representative seed colours for taxa, determine the phylogenetic signal of seed colour, and build objective classification categories for shape with high levels of visual interpretability.
4. Our approach increases productivity and allows for large-scale analyses that would otherwise be unfeasible. *Traitor* enables the acquisition of data that are readily comparable across different taxa, opening new avenues to explore functional relevance of morphological traits and to advance on new tools for seed identification.

This is an open access article under the terms of the [Creative Commons Attribution](https://creativecommons.org/licenses/by/4.0/) License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2023 The Authors. *Methods in Ecology and Evolution* published by John Wiley & Sons Ltd on behalf of British Ecological Society.

KEYWORDS

diaspores, high-throughput phenotyping, image segmentation, interpretability, morphological description, seed morphology, seed traits, trait measurement

1 | INTRODUCTION

The extraction of morphological traits from biological objects, such as seeds, is a highly specialised task that has historically relied on manual quantification and classification of morphological attributes. Manual quantitative measurements of seed size and surface structure (Kleyer et al., 2008) are time-consuming, limiting the number of species and replicates that can be measured. Contrastingly, assigning labels to seed traits based on visual perception can be a relatively quick process, but lacks sufficient objectivity as subtle differences in colour and shape among seeds require refined categories to separate taxa (Choi et al., 2012; Shimai, 2022) and leads to frequent inconsistencies in classification. These limitations have hindered the advancement of seed identification tools and the compilation of standardised seed morphological traits at global scale (Saatkamp et al., 2019).

Image analysis methods offer an alternative approach for rapid and objective characterisation of biological objects. The first step for image processing is the segmentation of objects from the background, which is critical for the accuracy of all measurements. Well-established methods, such as thresholding (Loddo et al., 2022; Olivoto, 2022) and deep learning (Ott & Lautenschlager, 2022; Schwartz & Alfaro, 2021), have been adapted for biologists with little or no coding experience, but their practicality is restricted when working with numerous species and structures. Thresholding can be inefficient when the seeds being measured are of a variety of sizes, surface structures and colours, causing optimal parameters to change, and even fail when seeds are either too small, glossy or have protruding structures. Deep learning-based segmentation requires a large amount of annotated data to train the network, which is limited by data availability and poses computational and time costs.

After segmenting the image, the next step involves extracting traits from each detected seed. While some applications for plant scientists are available as flexible tools to perform various morphological analyses, these require prior knowledge of Python (Gehan et al., 2017; Lürig, 2022). In contrast, most automatic seed measurement software provide users with only a few morphometric measurements (e.g. Tanabata et al., 2012; Zhu et al., 2021). Although there are options to automatically extract a vast array of features, these are mainly intended for seed identification classifiers (Loddo et al., 2022), rather than providing data that can be interpreted and visualised. Thus, current available tools do not adequately support users who work with a diverse range of taxa to enhance the characterisation of seed morphology and build standardised category systems for colour and shape.

Here, we propose an image analysis approach to enhance time-efficiency and objectivity of morphological trait extraction, with the capability to automatically handle seeds of different colours,

shapes, and structures. We describe a simple image acquisition and processing protocol and introduce *Traitor*, an open-source software available as a command-line interface (CLI), which automates the extraction of seed size, shape, and colour from images. Its name alludes to two ideas: a software that carries out trait extractions, and one that 'betrays' the seeds (and other biological objects) by giving away their valuable information to users. *Traitor* uses unsupervised segmentation to automate the separation of seeds from high-contrast backgrounds, eliminating the need for training, manual fine-tuning, or thresholding according to the particularities of each taxon/structure. The software is capable of processing hundreds of images of varied taxa simultaneously with just a few steps, allowing for large-scale analysis that would otherwise be unfeasible. It also provides means to easily verify the quality of image segmentation. We give detailed instructions for seed trait extraction, evaluate the accuracy of the protocol using images of 1228 taxa and provide case studies demonstrating how data extracted by *Traitor* can be used in ecological and evolutionary studies.

2 | MATERIALS AND METHODS

The workflow for seed trait extraction consists of sample preparation, image acquisition, image processing and trait measurement with *Traitor* software (Figure 1).

2.1 | Sample preparation

Seeds should be separated from debris, especially same-sized debris that cannot be numerically filtered. Appendages can be included with the following exceptions: (1) hairy appendages (e.g. pappus) as the static electricity on the scanner glass often displaces light, hairy structures, hindering image acquisition; (2) extremely fine structures (e.g. long hairs) because the algorithm for image segmentation may not handle fine structures blending into the background; (3) bent elongated appendages (e.g. bent awns), which may not be properly aligned resulting in incorrect size measurements. We recommend prior tests and adjustments for these three cases.

2.2 | Image acquisition

Images of seeds should be acquired observing the following requirements: (1) the background should be homogeneous and distinct in colour from the seeds; (2) lighting must be uniform and conditions should be the same across all images; (3) sharp shadows should be avoided, for example, by allowing some space between seeds and

background; (4) the scale must be known. The present protocol uses flatbed scanners covered with a frame approximately 10mm thick with a royal blue background to obtain the images due to the convenience of meeting these requirements (Figure 2; A HP Scanjet G4010 was used to build the dataset in Section 3). Seeds must be arranged without any contact or overlap, avoiding the edges of the scanner due to background shadows and the potential for lateral chromatic aberration (Matsuoka et al., 2012). Image resolution should be chosen based on the size of seeds: a higher resolution is needed to well-represent small seeds and irregular surfaces. All automatic correction functions provided by the scanner software should be disabled to ensure that the RGB values of the samples are not manipulated. It is recommended that resulting images are saved in TIFF or PNG formats, although the JPEG format is also acceptable

provided the images are acquired with a high spatial resolution and chrominance subsampling is minimised.

2.3 | Image processing

Colour measurements should be standardised to optimise their reproducibility and allow comparison between measurements obtained with different equipment and set ups (Stevens et al., 2007). Users can accomplish this by scanning a colour chart containing colour standards under the same settings as the seeds. The image of the colour chart is used to calculate a colour conversion matrix, based on a least squares fit in linear-RGB space, which should then be applied to images for optimal colour reproduction. A Spyder

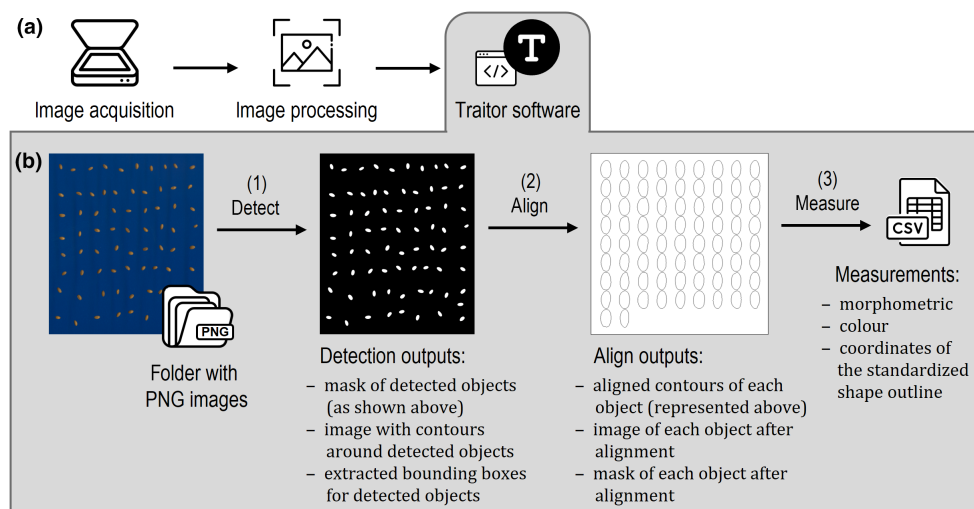


FIGURE 1 Workflow overview for trait extraction.

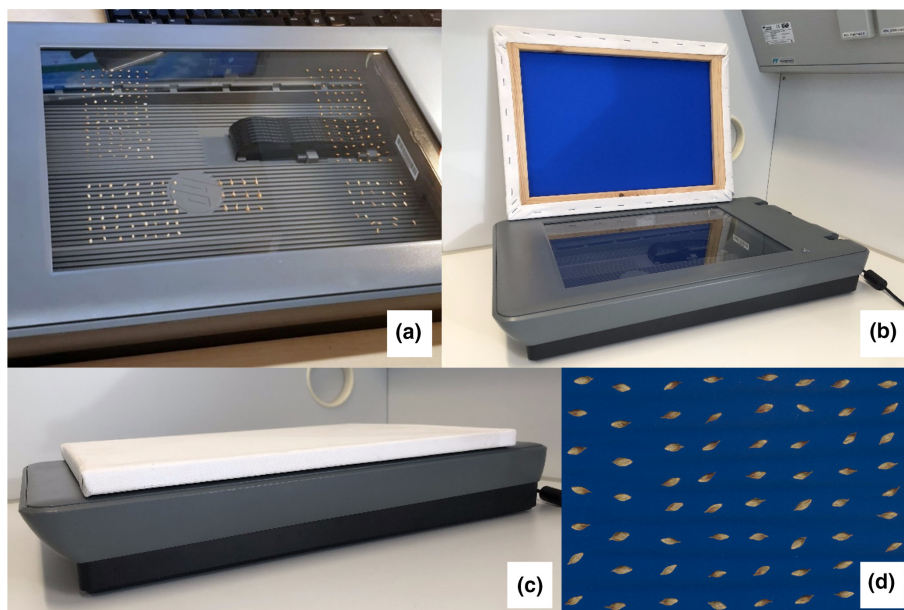


FIGURE 2 (a) Seeds arranged on the glass of the flatbed scanner. (b) The flatbed scanner and a wooden frame 10mm thick with a royal blue background (c), which is used as a lid. (d) Image of seeds after image processing.

Checkr® 24 card (Datacolor, NJ, USA) was used to build the dataset in Section 3 (code available at https://github.com/rdayrell/colour_calibration). Scanned images should be cropped to areas containing seeds, avoiding objects or shadows on the background. Save images in PNG format with no compression, to avoid quality loss.

2.4 | Software

Traitor is an image analysis software for automated extraction of measurements from images containing seeds on a high-contrast background. The software was built on Python open-source libraries, such as OpenCV and scikit-learn (Bradski, 2000; Pedregosa et al., 2011). It is capable of automatically processing multiple images simultaneously and measures all traits with only three commands (Figure 1b), which are explained in detail below. All functionalities are accessible via CLI.

2.4.1 | Extract

Traitor uses *k*-means clustering (Pedregosa et al., 2011) to separate seeds from a high-contrast background without the need for training, fine-tuning or thresholding. The algorithm works by iteratively grouping pixels into a predefined number of clusters based on their similarity and assigning each pixel to the cluster with the closest centroid, effectively separating the foreground from the background. Once PNG files are placed in a directory, the 'extract' command segments the image, detecting each object. A binary mask of the image with the detected outline of the detected objects (e.g. seeds) is created and serves as input for the 'align' command. *Traitor* can also generate two optional outputs for convenience: (1) A bounding box can be created with or without background removal for every detected object within the image, which can then be used in machine learning approaches; (2) The contours of every detected object can be drawn on the image, allowing users to easily assess the quality of the segmentation. All output files are created in sub-directories (one for each image) within the chosen output directory.

If several images do not exhibit a satisfactory result in this step, users should optimise image acquisition observing the requirements for *Traitor*'s optimal use (Section 2.2). If only a few problematic cases occur, users should first retry the 'extract' command with these images: as *k*-means relies on a random initialisation, additional runs can lead to satisfactory results. As a last resort to correct for specific problems, it is possible to edit the mask or the images to correct for minor problems (e.g. delete large debris from the image).

2.4.2 | Align

Next, the software aligns images. The first contour is aligned in an upright position according to its minimal bounding ellipse. All

other contours are then rotated to best match the reference contour using Procrustes analysis, which finds the optimal translation, scaling and rotation that minimises the sum of squared differences between corresponding points on the contours. The Procrustes analysis is not applied to the raw contours, but instead to a low-complexity approximation of the shapes generated using elliptic Fourier analysis. This complexity reduction was found to increase the algorithm's ability to successfully align damaged or irregularly shaped objects.

The 'align' command creates three outputs in separate directories: (1) 'contours' contains csv files with xy coordinates of the aligned contour of each detected object in the image—these files are used by the 'measure' command to obtain morphometric traits and standardised shape values; (2) 'extractions' contains cropped images of aligned objects; (3) 'masks' contains binary masks of the aligned objects, which are used to extract colour measurements.

2.4.3 | Measure

The 'measure' command uses the output of the 'align' command to obtain several measurements relevant for seed characterisation (Table 1; Appendix S1). The aligned contours are used to calculate traditional morphometric descriptors. Contours are also used to extract standardised shape coordinates, which can then be used to build objective classification categories with high levels of visual interpretability (e.g. Victorino & Gómez, 2019). Cropped images and masks are used to obtain two types of colorimetric measurements: sRGB values suitable for human recognition purposes; and linearised sRGB values, which are correlated with reflectance values in the three broad-band parts of the spectrum (longwave, shortwave, mediumwave) and thus, useful for studies independent of any particular animal visual system (Stevens et al., 2007). Seed count is not directly measured but can be easily obtained by row counts.

2.4.4 | Installation and usage

Traitor is implemented in Python (compatible with versions 3.7 to 3.10) and should be installed and used on terminal (Apple and Linux) or Command Prompt/Powershell (Windows). The software and all its dependencies can be installed using the PIP package installer. All commands provide a help option to list available parameters. *Traitor* source code, documentation and tutorial are available on GitHub (<https://github.com/TankredO/traitor>).

3 | VALIDATION AND CASE STUDIES

We validate the method and demonstrate the application of the extracted traits using images from DiasMorph, a dataset containing images of Central European seeds and diaspores (Dayrell, Begemann,

TABLE 1 Measurements extracted by *Traitor* (see Appendix S1 for details).

Traitor output	Description
Morphometric measurements	
Length	The longest dimension of the seed
Width	The widest axis perpendicular to the length axis
Aspect ratio	Proportional relationship between width and length
Area	Area of seed contour
Perimeter	Perimeter of seed contour in pixels
Surface structure	Shape descriptor that describes the roughness of an object surface
Solidity	Shape descriptor that describes the overall concavity of a seed
Circularity	Shape descriptor that describes the degree of similarity to a perfect circle
Colour measurements	
sRGB	Median and mean pixel value in sRGB
Linearised sRGB	Median and mean pixel value in linear sRGB
Dominant colours	Median pixel value in sRGB of most dominant colours. The number of colours is chosen by the user and must be a power of two Fraction of pixels assigned to each of the dominant colours are given
Shape	
Standardised shape	50 coordinates of the shape outline standardised for size invariance

et al., 2023). Images were acquired and processed and used for trait extraction following the protocol outlined in Section 2. The dataset contains seeds of a wide range of taxa with lengths varying from 0.33 to 30.2 mm, in a variety of shapes, colours, and appendages. All analyses using the extracted seed traits were performed in R (R Development Core Team, 2022) and are available on Github (https://github.com/rdayrell/Traitor_analyses).

3.1 | Validation of size measurements

To validate the results, we used images of 91,667 seeds from 1228 taxa (≥ 5 seeds per taxa; 89 plant families) from the DiasMorph dataset. We compared the average length and width values extracted by *Traitor* with average manual measurements obtained from seeds of the same collection (Supporting Information); a one-to-one correspondence between the seeds could not be established. For manual measurements, five seeds from each taxon were measured with the aid of a stereo microscope (Stemi SV 11; Carl Zeiss Jena GmbH, Germany), following standardised protocol (Kleyer et al., 2008). The agreement between the measurements obtained by the two

methods was assessed by calculating Lin's concordance correlation coefficient (ρ_c), an index of how well a new measurement reproduces a standard measurement that ranges between 1, perfect concordance, and -1, perfect discordance (Lin, 1989). For this, we used the 'CCC' function implemented in the 'DescTools' package (Signorell et al., 2022).

The ρ_c for length and width measurements were 0.979 (95% CI [0.976–0.981]; Figure 3a,b) and 0.984 ([0.983–0.986]; Figure 3c,d), respectively. This shows a high agreement with the reference values from *Traitor's* measurements. We manually checked the outlines and alignment outputs of the measurements with less than 95% agreement between both measurements and detected no issues with *Traitor's* extract and align outputs. As we have not obtained the same measurement from the exact same seeds, the detection of larger differences in a few cases can be attributed to measurements being carried out on a less representative sample of seeds (with a size bias) or different structures (e.g. seeds rather than diaspores).

3.2 | Colour measurements for human recognition

We used *Traitor's* output for taxa of the Rosaceae family within the DiasMorph dataset to obtain representative colours for taxa which can assist on tasks related to colour description for human recognition (see Appendix S2 for a detailed method description). Median sRGB values were used to describe colour variation among taxa: channel-wise median of seeds' median colours was calculated for each taxon; seeds' median colours were used in principal component analysis (PCA) for dimensionality reduction, and the colour corresponding to the lowest and the highest PC1 value were considered as minimum and maximum colour values for the taxon (Figure 4). The channel-wise medians of seeds' dominant colours were used to describe colour variation in seeds of each taxon (Figure 4; Appendix S3).

3.3 | Colour measurements for ecological and evolutionary investigations

Linearised sRGB values exhibit a linear response to changes in light intensity and are a convenient option to obtain colour measurements for ecological and evolutionary studies (Stevens et al., 2007). In this case study, we sampled colour traits of taxa in the Asteraceae family within the DiasMorph dataset and assessed whether seed colour is correlated to phylogenetic relatedness (see Appendix S2 for a detailed method description). Briefly, we calculated the median value of linearised sRGB for each taxon and subsequently performed a PCA for dimensionality reduction. Pagel's λ was used to estimate phylogenetic signal in PC1 scores, which explained 93.6% of the variation. λ can vary from zero, no correlation between species, and 1, species' traits are distributed as expected under Brownian motion (Pagel, 1999). The λ for PC1 was 0.84, and the likelihood ratio test rejected the hypothesis of lack of correlation among species ($p < 0.0001$). The result is consistent with some taxonomic groups

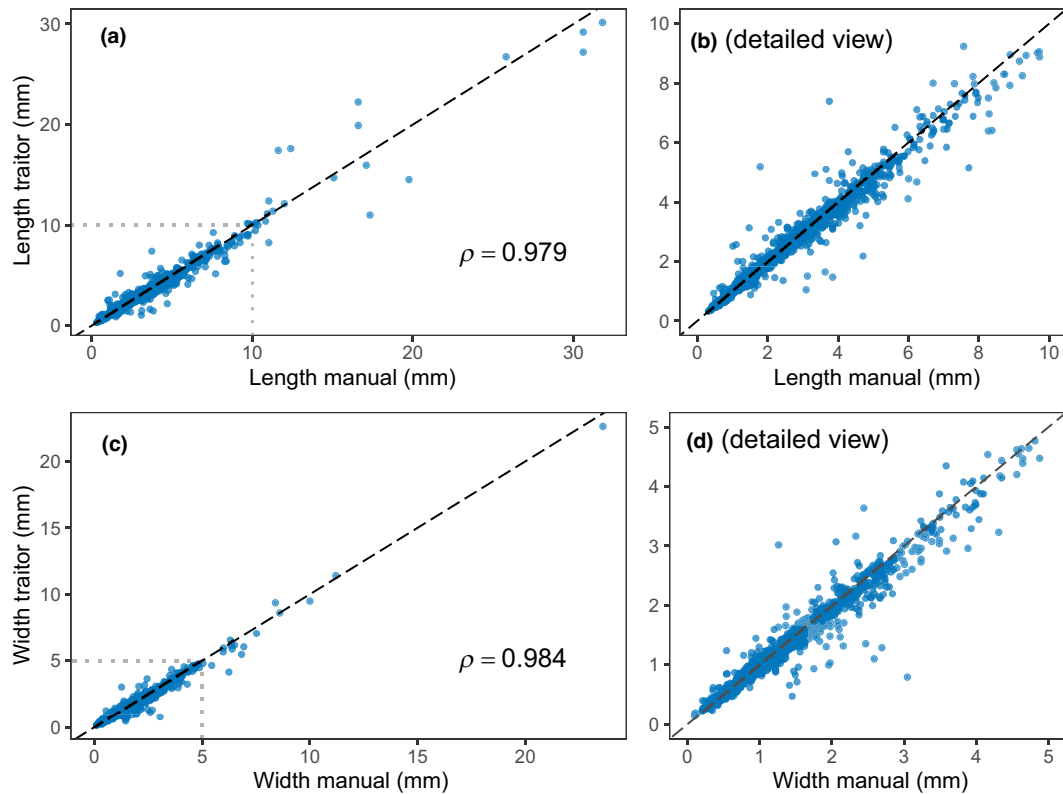


FIGURE 3 Relationship between manual and *Traitor*'s measurements: (a) Length and (c) width, with (b, d) respective zoomed-in plots. Dots are mean values for each taxon ($n = 1228$). The dashed line represents the perfect correspondence between manual and *Traitor*'s measurements.

having a greater propensity towards darker colours, and with multiple evolutionary transitions of colour traits (Figure 5).

3.4 | Objective categorisation of seed shapes

We used records of *Carex* species within the DiasMorph dataset to demonstrate how the shape outline standardised for size invariance (*Traitor*'s output), can be used for the construction of objective and interpretable categories. Briefly, outlines were aligned, represented by an Elliptic Fourier transform as quantitative variables (harmonics), which were then used in PCA for dimensionality reduction. Hierarchical cluster analysis was then used to determine the shape categories (see Appendix S2 for a detailed method description). We divided shapes into eight groups to exemplify how this approach can improve communication of subtle differences. The shape closest to the centre of each cluster is the representative shape of the cluster (Figure 6b; see Appendix S4 for outline of all seeds within each cluster). We overlaid 95% data ellipses for seeds of six species to visualise differences in intraspecific variability in seed shape.

PC1 explained 70.7% of the variance of the set of harmonics, separating seeds by aspect ratio; PC2 explained 14.3% of the variance, separating seeds that are relatively symmetrical along their width from seeds that widen below the centre while tapering above

the centre (Figure 6a,b). Intraspecific variability can greatly differ among species as evidenced by contrasting data dispersion of species: for example *Carex paucifolia* seeds were concentrated in only one cluster, while *Carex atrata* subsp. *aterrima* seeds were found in five different clusters (Figure 6c).

4 | DISCUSSION AND CONCLUDING REMARKS

The image acquisition and processing protocol and *Traitor*, the open-source software, presented in this study were designed to drastically speed up the extraction of seed morphological traits. The method aims to support scientists and practitioners who work with a wide range of taxa to obtain objective and interpretable measurements in a time and labour-saving manner. *Traitor*'s main strengths are its high level of automation in the image segmentation process, and its ability to produce highly accurate and interpretable measurements regardless of the morphological attributes of the measured object. Additionally, *Traitor* allows users to easily inspect the quality of image segmentation outputs. Therefore, our approach provides a more convenient and automated alternative to extract interpretable traits from seeds with diverse morphological attributes than user-friendly open-source solutions, such as ImageJ plugins (Loddo et al., 2022), image processing pipelines in

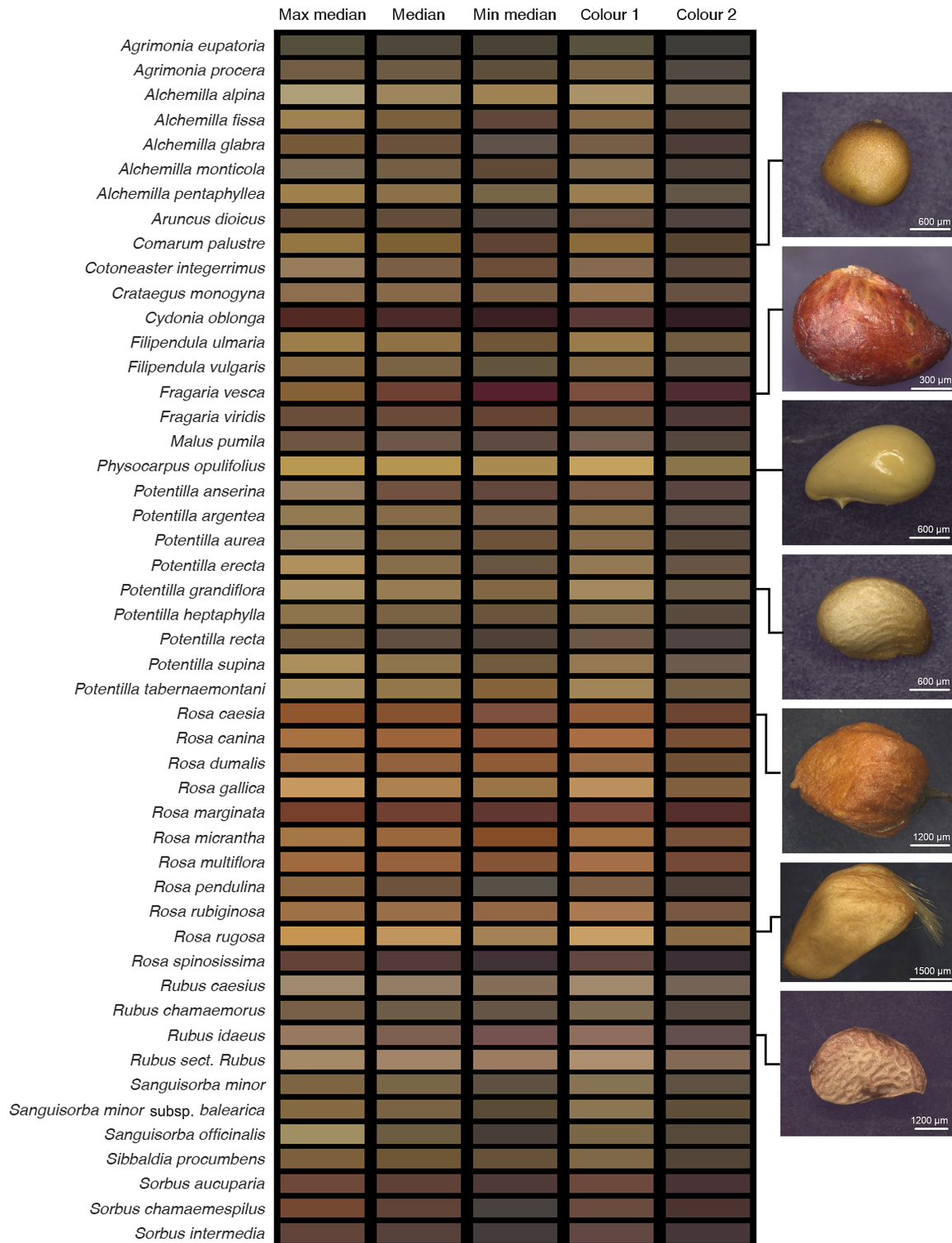


FIGURE 4 Representative colours of taxa from the Rosaceae family. 'Median': channel-wise median. 'Min median' and 'max median': extremes of median colour range of each taxon based on a PCA. 'Colour 1' and 'colour 2': channel-wise median of the two most dominant colours in the seeds. Hex codes in Supporting Information. Representative photographs of seeds obtained using an optical microscope (Zeiss Axio Imager.M2m). Photographs: Maxi Bleicher and Irina Weinberger.

Python (Gehan et al., 2017) and deep-learning based models (Ott & Lautenschlager, 2022).

The method described in this work has the potential to measure other biological objects, such as leaves, and *Traitor* can process

images obtained using different acquisition procedures, as long as the requirements outlined in Section 2.2 are met. However, it is important to note that the method's performance may be affected by the complexity of the structures being measured, and

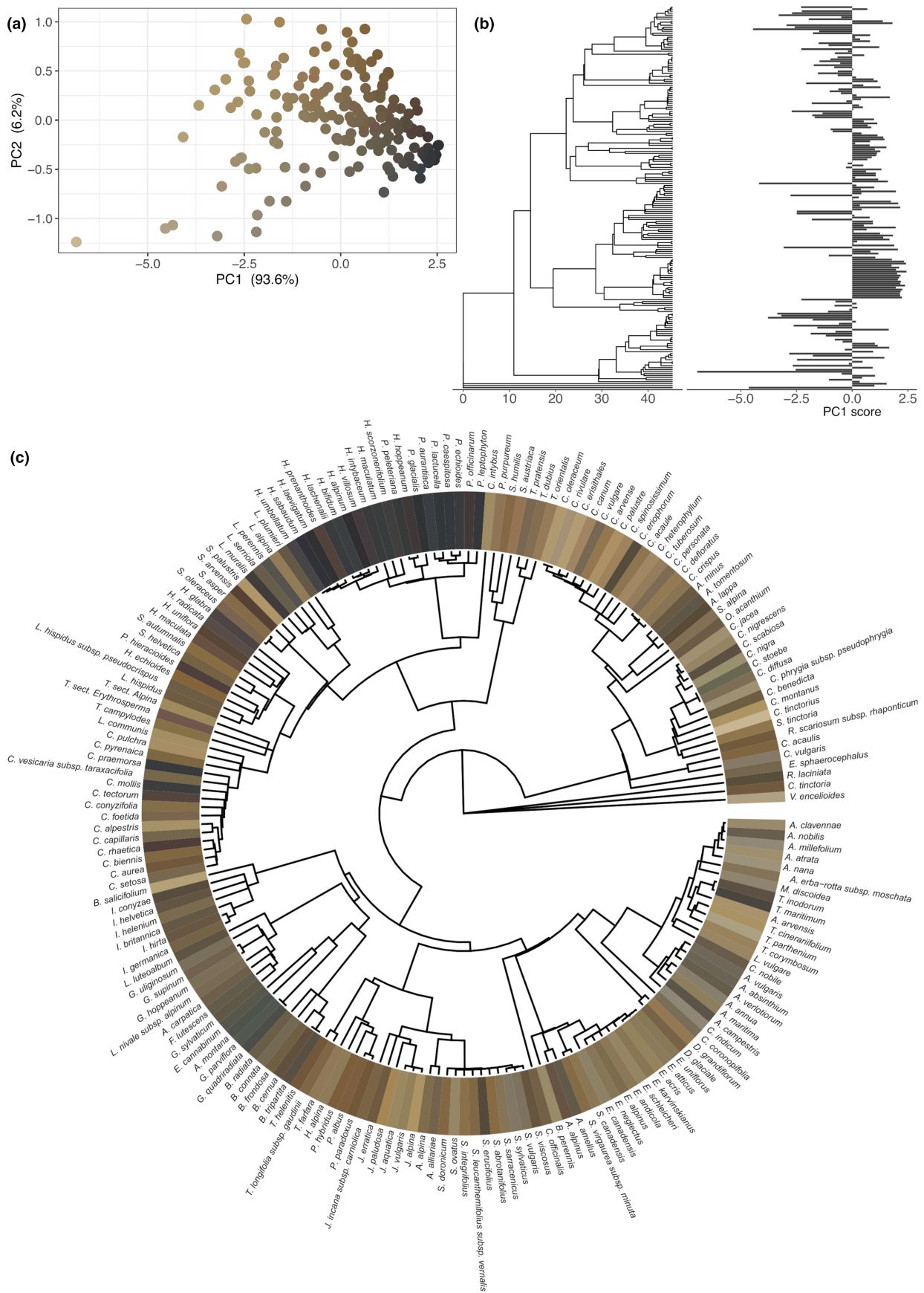


FIGURE 5 Seed colour of taxa in the Asteraceae family. (a) PCA on median value of linearised sRGB for dimensionality reduction. (b) PC1 scores mapped onto the phylogeny of sampled taxa. (c) Seed colour in sRGB mapped onto the phylogeny of sampled taxa for visualisation purposes.

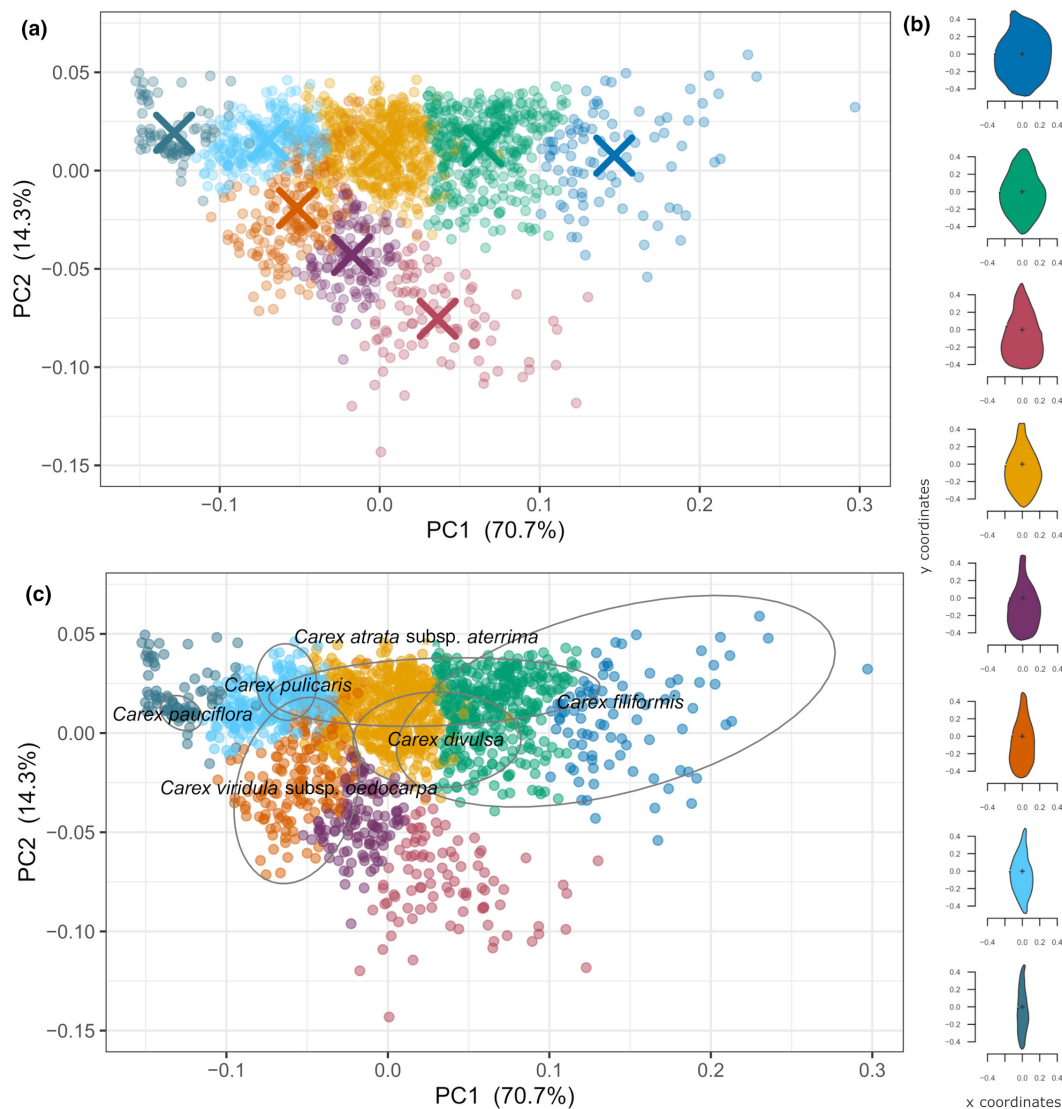


FIGURE 6 Categorisation of standardised seed shapes of *Carex* spp. ($n=56$). (a, c) PCA showing the two main axes of variability in seed shape. (a) 'X' symbol denotes the centre of the cluster with the same colour. (b) The shapes closest to the centre of each cluster plotted in corresponding colours. (c) 95% data ellipses of six *Carex* species in the shape space.

the interpretability of the extracted data can also be compromised. Users should carefully evaluate the method's suitability for other applications and set-ups.

Traitor's output provides means for constructing category systems for seed shape and colour with high levels of visual interpretability, offering opportunities to improve the communication of seed morphology and advance in new identification tools. The approach can open new avenues in the field of seed ecology and evolution and increase data availability of seed morphology which are comparable across taxa.

AUTHOR CONTRIBUTIONS

Roberta Dayrell designed the study, conducted case studies, and wrote the manuscript. Tankred Ott programmed the software. Tom

Horrocks programmed the colour calibration and advised on the use of colour spaces. Peter Poschlod provided his seed herbarium, conceived the project and obtained the funding. All authors contributed critically to the drafts.

ACKNOWLEDGEMENTS

This research was funded by the European Regional Development Fund—ERDF Project 343 within the Target ETC Program Free State of Bavaria—Czech Republic 2014-2020 (INTERREG V). Inge Lauer carried out manual measurements of seeds. Open Access funding enabled and organized by Projekt DEAL.

CONFLICT OF INTEREST STATEMENT

The authors declare no conflict.

PEER REVIEW

The peer review history for this article is available at <https://www.webofscience.com/api/gateway/wos/peer-review/10.1111/2041-210X.14127>.

DATA AVAILABILITY STATEMENT

All codes used in this publication are available at GitHub and deposited at Zedono as follows: TraitOr—<https://github.com/TankredO/traitor> (Ott, 2023); colour calibration—https://github.com/rdayrell/colour_calibration (Horrocks, 2023); validation and case studies analyses—https://github.com/rdayrell/TraitOr_analyses (Dayrell, 2023). Supporting Information, including manual seed measurements dataset, hex codes for Rosaceae seeds, and Appendices S1–S4, available via the Figshare Repository <https://doi.org/10.6084/m9.figshare.22272535> (Dayrell, Ott, et al., 2023). DiasMorph image dataset available via the Figshare Repository <https://doi.org/10.6084/m9.figshare.21206507> (Dayrell, Begemann, et al., 2023).

ORCID

Roberta L. C. Dayrell  <https://orcid.org/0000-0002-4770-9100>

Tankred Ott  <https://orcid.org/0000-0001-6748-0510>

Tom Horrocks  <https://orcid.org/0000-0003-0835-6151>

Peter Poschlod  <https://orcid.org/0000-0003-4473-7656>

REFERENCES

- Bradski, G. (2000). The OpenCV library. *Dr. Dobb's Journal of Software Tools*, 25(11), 120–126.
- Choi, B. K., Duretto, M. F., & Hong, S. P. (2012). Comparative seed morphology of *Boronia* and related genera (Boroniinae: Rutaceae) and its systematic implications. *Nordic Journal of Botany*, 30(2), 241–256. <https://doi.org/10.1111/j.1756-1051.2011.01251.x>
- Dayrell, R. L. C. (2023). rdayrell/TraitOr_analyses: Code from: Dayrell RLC, Ott T, Horrocks T, Poschlod P (2023) Automated extraction of seed morphological traits from images. *Zenodo*, <https://doi.org/10.5281/zenodo.7877500>
- Dayrell, R. L. C., Begemann, L., Ott, T., & Poschlod, P. (2023). DiasMorph: A dataset of morphological traits and images of Central European diaspores. *Authorea* (Preprint). <https://doi.org/10.22541/au.167935473.32803184/v2>
- Dayrell, R. L. C., Ott, T., Horrocks, T., & Poschlod, P. (2023). Supplementary material from “Automated extraction of seed morphological traits from images”. *figshare*, <https://doi.org/10.6084/m9.figshare.22272535>
- Gehan, M. A., Fahlgren, N., Abbasi, A., Berry, J. C., Callen, S. T., Chavez, L., Doust, A. N., Feldman, M. J., Gilbert, K. B., Hodge, J. G., Hoyer, J. S., Lin, A., Liu, S., Lizárraga, C., Lorence, A., Miller, M., Platon, E., Tessman, M., & Sax, T. (2017). PlantCV v2: Image analysis software for high-throughput plant phenotyping. *PeerJ*, 5, e4088. <https://doi.org/10.7717/peerj.4088>
- Horrocks, T. (2023). Rdayrell/colour_calibration: Colour calibration code from: Dayrell RLC, Ott T, Horrocks T, Poschlod P (2023) Automated extraction of seed morphological traits from images. *Zenodo*, <https://doi.org/10.5281/zenodo.7877540>
- 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., Klotz, S., Rusch, G. M., Hermy, M., Adriaens, D., Boedeltje, G., Bossuyt, B., Dannemann, A., Endels, P., Götzenberger, L., ... Peco, B. (2008). The LEDA Traitbase: A database of life-history traits of the Northwest European flora. *Journal of Ecology*, 96(6), 1266–1274. <https://doi.org/10.1111/j.1365-2745.2008.01430.x>
- Lin, L. I.-K. (1989). A concordance correlation coefficient to evaluate reproducibility. *Biometrics*, 45(1), 255. <https://doi.org/10.2307/2532051>
- Loddo, A., Di Ruberto, C., Vale, A. M. P. G., Uccesu, M., Soares, J. M., & Bacchetta, G. (2022). An effective and friendly tool for seed image analysis. *The Visual Computer*, 39, 335–352. <https://doi.org/10.1007/s00371-021-02333-w>
- Lürig, M. D. (2022). Phenotype: A phenotyping pipeline for Python. *Methods in Ecology and Evolution*, 13(3), 569–576. <https://doi.org/10.1111/2041-210X.13771>
- Matsuoka, R., Asonuma, K., Takahashi, G., Danjo, T., & Hirana, K. (2012). Evaluation of correction methods of chromatic aberration in digital camera images. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 1–3, 49–55. <https://doi.org/10.5194/isprsannals-1-3-49-2012>
- Olivoto, T. (2022). Lights, camera, pliman! An R package for plant image analysis. *Methods in Ecology and Evolution*, 13(4), 789–798. <https://doi.org/10.1111/2041-210X.13803>
- Ott, T. (2023). TankredO/traitor: 0.1.0. *Zenodo*, <https://doi.org/10.5281/zenodo.7876319>
- Ott, T., & Lautenschlager, U. (2022). GinJinn2: Object detection and segmentation for ecology and evolution. *Methods in Ecology and Evolution*, 13(3), 603–610. <https://doi.org/10.1111/2041-210X.13787>
- Pagel, M. (1999). Inferring the historical patterns of biological evolution. *Nature*, 401(6756), 877–884. <https://doi.org/10.1038/44766>
- Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., & Duchesnay, É. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12(85), 2825–2830. <http://jmlr.org/papers/v12/pedregosa11a.html>
- R Development Core Team. (2022). R: A language and environment for statistical computing (4.2.1). Foundation for Statistical Computing. <http://www.rproject.org>
- Saatkamp, A., Cochrane, A., Commander, L., Guja, L. K., Jimenez-Alfaro, B., Larson, J., Nicotra, A., Poschlod, P., Silveira, F. A. O., Cross, A. T., Dalziell, E. L., Dickie, J., Erickson, T. E., Fidelis, A., Fuchs, A., Golos, P. J., Hope, M., Lewandowski, W., Merritt, D. J., ... Walck, J. L. (2019). A research agenda for seed-trait functional ecology. *New Phytologist*, 221(4), 1764–1775. <https://doi.org/10.1111/nph.15502>
- Schwartz, S. T., & Alfaro, M. E. (2021). Sashimi: A toolkit for facilitating high-throughput organismal image segmentation using deep learning. *Methods in Ecology and Evolution*, 12(12), 2341–2354. <https://doi.org/10.1111/2041-210X.13712>
- Shimai, H. (2022). Seed shape supports phylogenetic relationships and adaptive radiation in the genus *Pinguicula* (Lentibulariaceae). *Botanical Journal of the Linnean Society*, 198(3), 342–359. <https://doi.org/10.1093/botlinnean/boab058>
- Signorell, A., Aho, K., Alfons, A., Anderegg, N., Aragon, T., Arachchige, C., Arppe, A., Baddeley, A., Barton, K., Bolker, B., Borchers, H. W., Caeiro, F., Champely, S., Chessel, D., Chhay, L., Cooper, N., Cummins, C., Dewey, M., Doran, H. C., ... Zeileis, A. (2022). *DescTools: Tools for descriptive statistics*. R package version 0.99.47.
- Stevens, M., Párraga, C. A., Cuthill, I. C., Partridge, J. C., & Trosianko, T. S. (2007). Using digital photography to study animal coloration. *Biological Journal of the Linnean Society*, 90(2), 211–237. <https://doi.org/10.1111/j.1095-8312.2007.00725.x>
- Tanabata, T., Shibaya, T., Hori, K., Ebana, K., & Yano, M. (2012). SmartGrain: High-throughput phenotyping software for measuring seed shape through image analysis. *Plant Physiology*, 160(4), 1871–1880. <https://doi.org/10.1104/pp.112.205120>

- Victorino, J., & Gómez, F. (2019). Contour analysis for interpretable leaf shape category discovery. *Plant Methods*, 15(1), 112. <https://doi.org/10.1186/s13007-019-0497-6>
- Zhu, F., Paul, P., Hussain, W., Wallman, K., Dhatt, B. K., Sandhu, J., Irvin, L., Morota, G., Yu, H., & Walia, H. (2021). SeedExtractor: An open-source GUI for seed image analysis. *Frontiers in Plant Science*, 11(February), 1–10. <https://doi.org/10.3389/fpls.2020.581546>

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

Appendix S1: Table containing detailed information on measurements extracted by Traitor.

Appendix S2: Detailed methods for Sections 3.2, 3.3, and 3.4.

Appendix S3: Figure showing colour variation within and among seeds of *Fragaria vesca*.

Appendix S4: Figure of categorisation of standardised seed outlines of *Carex* spp. ($n=56$) within the DiasMorph dataset, depicting the representative shape of clusters and the outlines of all seeds within the cluster.

How to cite this article: Dayrell, R. L. C., Ott, T., Horrocks, T., & Poschlod, P. (2023). Automated extraction of seed morphological traits from images. *Methods in Ecology and Evolution*, 14, 1708–1718. <https://doi.org/10.1111/2041-210X.14127>