



OPEN

DATA DESCRIPTOR

Reference bioimaging to assess the phenotypic trait diversity of bryophytes within the family Scapaniaceae

Kristian Peters^{1,2,3} & Birgitta König-Ries^{1,4,5}

Macro- and microscopic images of organisms are pivotal in biodiversity research. Despite that bioimages have manifold applications such as assessing the diversity of form and function, FAIR bioimaging data in the context of biodiversity are still very scarce, especially for difficult taxonomic groups such as bryophytes. Here, we present a high-quality reference dataset containing macroscopic and bright-field microscopic images documenting various phenotypic characters of the species belonging to the liverwort family of Scapaniaceae occurring in Europe. To encourage data reuse in biodiversity and adjacent research areas, we annotated the imaging data with machine-actionable metadata using community-accepted semantics. Furthermore, raw imaging data are retained and any contextual image processing like multi-focus image fusion and stitching were documented to foster good scientific practices through source tracking and provenance. The information contained in the raw images are also of particular interest for machine learning and image segmentation used in bioinformatics and computational ecology. We expect that this richly annotated reference dataset will encourage future studies to follow our principles.

Background & Summary

In biodiversity, organisms are studied with the aim to record their diversity at the genetic, metabolic, physiological, morphological, or the ecosystem level. Despite the fact that bioimaging techniques such as macro- and microscopy are prominently used, FAIR bioimaging data in the context of biodiversity are still very scarce^{1–3}. This is especially the case for taxonomically difficult and underrepresented groups such as bryophytes. Currently, there are approx. 24'000 species of bryophytes known to science⁴. Unlike vascular plants, bryophytes lack well-differentiated organs that protect them from environmental exposures and pathogens. As a result, phenotypes are often cryptic and difficult to identify visually as bryophytes have developed unique specialised metabolisms and cell structures such as oil bodies^{5,6}.

The highly diverse family of Scapaniaceae contains 48 taxa in Europe⁷ and is an ecologically important group regarding environmental adaptations⁸, the biochemistry of terpenoid natural products and other chemical structures^{9–12}, the metabolism of pollutants and heavy metals^{13,14}, and phylogenetics^{15–18}. Generally, there is a considerable lack of described traits in bryophytes¹⁹ and especially in liverworts such as Scapaniaceae, phenotypic traits to assess the diversity of form and function are understudied²⁰.

Bioimaging data in the field of biodiversity is of high relevance as they allow to assess the phenotypic diversity through an analysis and assessment of images^{2,21,22}. In the form of measurable phenotypic traits, biological images are the groundwork of many ecological studies^{3,20,21,23,24}. Phenotypisation through recording images of anatomical and morphological characters allows qualitative and quantitative measurements of molecular structures relating to genetics, molecular pathways and biotechnology^{25–27}. Bioimages have also gained a lot of interest

¹German Centre for Integrative Biodiversity Research (iDiv) Halle-Jena-Leipzig, Puschstraße 4, 04103, Leipzig, Germany. ²Institute of Biology/Geobotany and Botanical Garden, Martin Luther University Halle-Wittenberg, Am Kirchtor 1, 06108, Halle (Saale), Germany. ³Bioinformatics and Scientific Data, Leibniz Institute of Plant Biochemistry, Weinberg 3, 06120, Halle (Saale), Germany. ⁴Heinz-Nixdorf Chair for Distributed Information Systems, Friedrich Schiller University, Jena, Germany. ⁵Michael Stifel Center Jena, Jena, Germany. [✉]e-mail: kpeters@ipb-halle.de

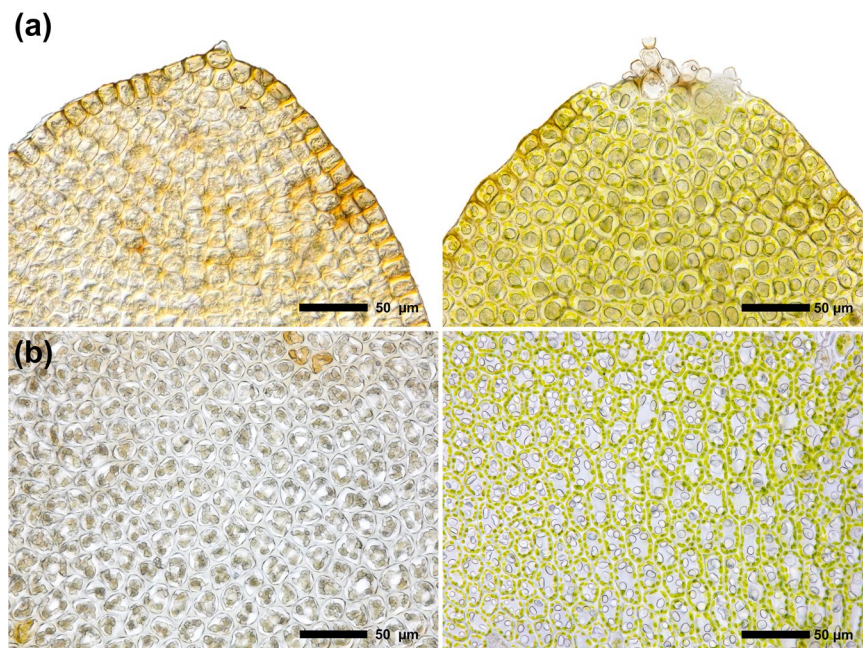


Fig. 1 Comparison of leaf cells of dried herbaria voucher specimens and fresh samples. Oil bodies are usually absent from dried specimens. **(a)** Cells in the apex of the antical lobe of *Scapania gymnostomophila* in a voucher specimen (left) and a fresh sample (right). Cells of this species produce one large brownish structured cellular oil body. **(b)** Cells in the centre of the postical lobe of *Scapania cuspiduligera* in a dried herbaria voucher specimen (left) and in a fresh specimen (right). Cells of this species usually produce 2–5 translucent oil bodies per cell.

in citizen sciences and in the digitization of natural history collections and digital herbaria^{28,29}. Furthermore, meta-synthesis methods, which synthesise disparate data sources spanning published case studies, have great potential to reveal context-dependencies within bioimaging research data³⁰.

For example, liverworts such as the species investigated herein produce cellular oil bodies that are visible under the microscope as little droplets of oil (Fig. 1). These oil bodies are often species-specific and are an important phenotypic character for species identification. Furthermore, as they contain many specialized metabolites such as fatty acids, terpenoids, or flavonoids they can provide a mechanistic link between molecular function and the phenotype^{4,5,31}. Images of oil bodies are of high interest as they often degrade within a few hours or days after sampling and are usually absent in dried herbaria material.

Macroscopy and microscopy are characterized by physical constraints resulting in diffraction and shallow depth of field^{22,32,33}. From a technical perspective, our data employs two major methods to significantly extend the depth of field and to increase the resolution of the composite images: image stitching (combining several images relative to the x- and y-axes of the visible accommodation to form an image with a larger frame)^{34,35} and multi-focus image fusion (merging multiple images at different focal planes of the z-axis in such a way that only regions in focus will contribute to the resulting image)^{36,37} (Figs. 2, 3). In this regard, the raw data also allows to be reused for combining image fusion with computational super-resolution^{22,38–41}.

Cloud infrastructural resources are able to execute computational workflows that combine data with computational analysis tools at a large scale^{42,43}. However, there is still a considerable lack of data containing machine-actionable metadata^{1,3,44–47}. To document provenance, ensure reproducibility and support reuse any raw and segmented image in this data set has been associated with a rich set of contextual and expressive metadata⁴⁸, documenting the phenotypic characters, and recording any digital image processing (i.e., increasing contrast, brightness, image fusion) (Fig. 3). The metadata has been annotated with community-accepted semantics that allow for machine-actionable data-mining and to create scientific workflow modules that produce segmented composite images automatically by reusing the instructions contained in our metadata^{20,43,46,47,49}.

In this Data Descriptor, we present the principles to generate reference images from raw microscopic bioimaging data and show how individual images are associated with technical and expressive metadata. Despite that we were able to associate our images with a rich set of metadata, we ascertain that there is still a lack of usable ontological terms and schemas in bioimaging with regard to documenting image processing and associating individual images with phenotypic characters^{3,46} (Table 1). Our high-resolution images allow for large prints and zooming into images to obtain critical details, which is particularly important for species identification and for computational image analyses, computer-assisted species recognition and identification^{50,51}. Despite that we have deposited the data to the two specialised imaging repositories BioStudies (containing raw and pre-processed images which enable direct use in, i.e., machine learning approaches in computational ecology) and Imaging Data Resource (containing pre-processed and fully segmented images to be rapidly reused by ecologists), we ascertain that there is still the need for connecting macro- and microscopic bioimaging data to

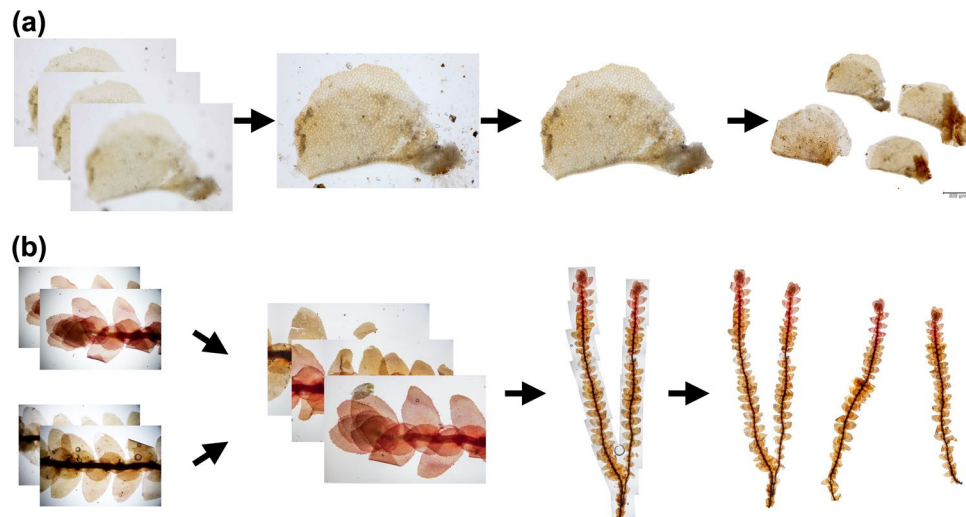


Fig. 2 Two exemplary processing workflows used in this study to create segmented images. **(a)** Example of multi-focus image fusion where (1) several images of one object of a leaf lobe are fused into a (2) composite image. The leaf lobe as shown in the composite image is then (3) segmented and the background removed. Several leaf lobe objects are then put onto the (4) final image and a microscopic scale is applied. **(b)** Example of image stitching where several fused images of the same object showing the ventral sides of the stature (habitus) of a plant are (2) arranged into segments and (3) stitched into a composite image with larger dimensions. Several of these stitched images are then put onto the (4) final image and a microscopic scale is applied.

biodiversity platforms^{3,52} such as iDigBio⁵³ and GBIF⁵⁴, or even the citizen scientists community-effort iNaturalist⁵⁵. Our reference data framework facilitates the further integration of bioimaging data into other research disciplines⁵⁶ and, thus, we want to inspire future data reuse and meta-synthesis in the fields of biodiversity and computational ecology.

Methods

Sample collection and biological material. Representative voucher specimens were received from different herbaria. Supplementary Table 1 lists all used voucher specimens and freshly collected samples that have been investigated in this study. Samples have been associated with taxonomic species identifiers (NCBI, GBIF, or *Open Tree of Life* identifiers, if available), the text on the specimen sleeves (collector, date and text on the envelopes) and the voucher specimen identifiers (the first letters either indicate the *Index Herbariorum* institution code⁵⁷, if available, or the name of private collection where the specimens were stored). Fresh samples of *Diplophyllum taxifolium*, *Scapania cuspiduligera*, *Scapania gymnostomophila* and *Scapania subalpina* were additionally investigated to depict oil bodies which are usually absent from herbaria specimens. Fresh samples were additionally collected at various sites, put into envelopes on-site, identified and photographed afterwards. Information regarding the date, site (including geographical coordinates), habitat, substrate and other further information were collected.

Microscopy and photographic equipment. For microscopy, a Zeiss Axio Scope.A1 HAL 100/HBO, 6x HD/DIC, M27, 10x/23 microscope with an achromatic-aplanatic 0.9 H D Ph DIC condenser was used with the objectives EC Plan Neofluar 2.5x/0.075 M27 (a = 8.8 mm), Plan-Apochromat 5x/0.16 M27 (a = 12.1 mm), Plan-Apochromat 10x/0.45 M27 (a = 2.1 mm), Plan-Apochromat 20x/0.8 M27 (a = 0.55 mm), and Plan-Apochromat 40x/0.95 Korr M27 (a = 0.25 mm) using the EC PN and the Fluor 40x/1.30 III and PA 40x/0.95 III filters for DIC. The conversion filter CB3 and the interference filter wideband green were used to improve digital reproduction of colors. Color balance was adjusted in the camera and during postprocessing of the images. For macroscopy and for preparing microscopy slides, a binocular microscope Zeiss Stemi 2000c was used (apochromatic Greenough system with a stereo angle of 11° and 100/100 switchover of camera and ocular viewing). The objectives Canon MP-E 65 mm 2.8 1-5x macro and Venus Optics Laowa 25 mm 2.5-5.0x ultra-macro for Canon EF and the Canon EF-RF adapter were used for stand-alone macroscopic images.

A full-frame, high-resolution camera (Canon EOS RP, 26 megapixel) was used to acquire digital images. It was adapted to the microscopes using binocular phototubes with sliding prism 30°/23 (Axio Scope.A1) and 100:0/0:100 reversed image (Stemi 2000c) using 60-T2 camera adapter for Canon EOS and Canon EF-RF adapter. The objectives Canon MP-E and Laowa 25 mm were adapted directly through the Canon EF-RF adapter.

Image processing. Figures 2 and 3 provide overviews on the image processing tasks that were performed. Images were recorded at different focal planes to construct images with extended depth of field using computational methods. This “focus stacking” approach was automatized for macroscopy by attaching the camera to a Cognisys StackShot macro rail fixed on a Novoflex macro stand, and for microscopy by adapting a Cognisys StackShot motor to the fine adjustment of the microscope using two cogged wheels, one small wheel (1 cm diameter) adapted on the motor and one large wheel (8.5 cm diameter) on the fine adjustment of the microscope. The

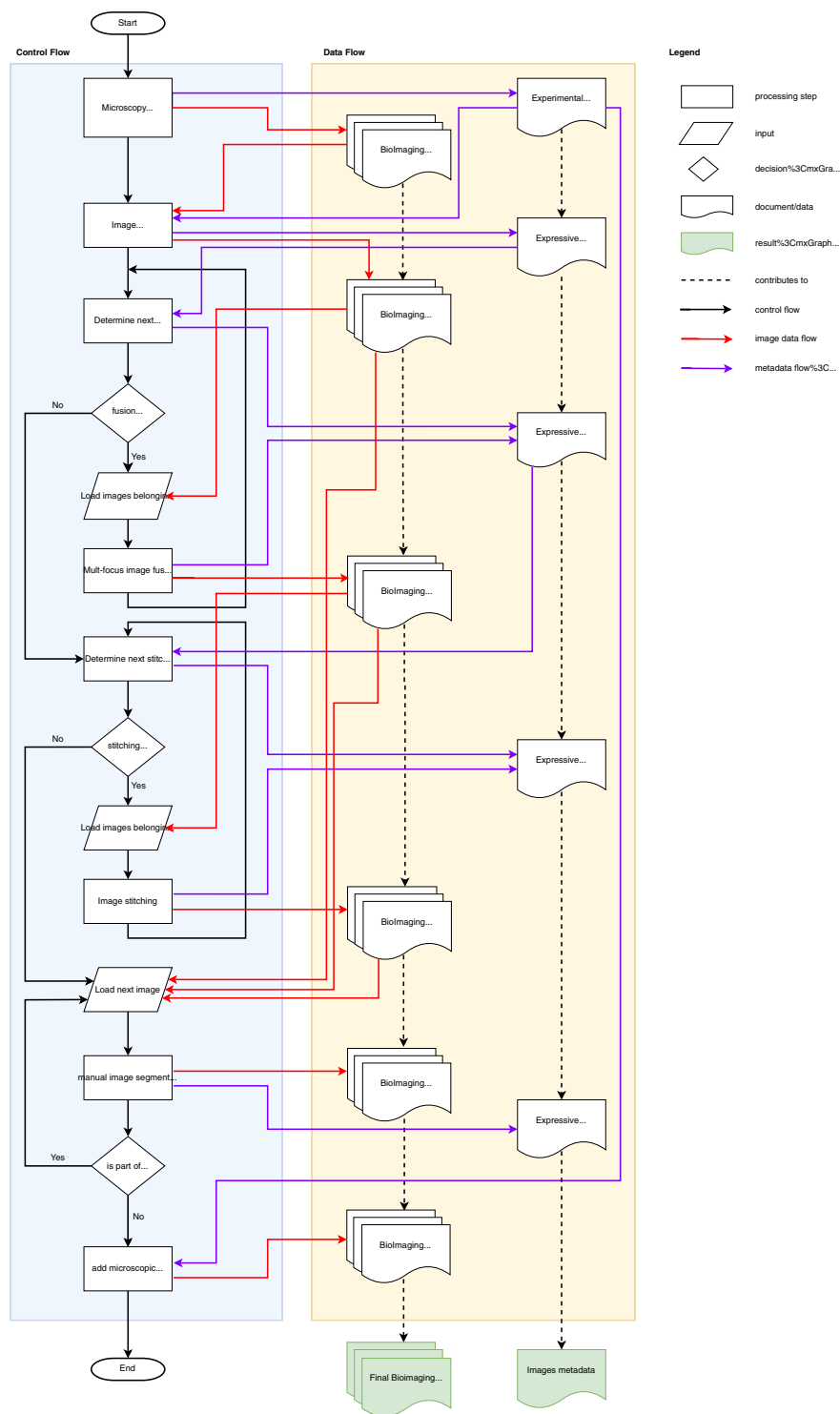


Fig. 3 Flow chart for the bioimage processing workflow for one species. The workflow starts with the microscopy experiment where raw bioimages are acquired for several biological objects. These raw images are pre-processed using image enhancement methods such as color balance, or exposure correction using experimental metadata and generating expressive metadata. These bioimaging data is then further processed using image fusion or image stitching methods where several images of the same object are fused or stitched together. The processed images are then manually segmented such as separating the object from the background, or putting segmented objects such as leaves onto one image. Finally, a microscopic scale is put onto the processed image using the metadata information. During each processing step, experimental data is recorded and annotated in the final image metadata. The flow chart was created using the draw.io web software tool.

Phenotypic character	Organisational type	Semantic term id
antheridia	cellular	PO_0025125
antical lobe	anatomical	
archegonia	cellular	PO_0025126
bordered margin	cellular	
capsule	morphological	UBERON_0003893
cells	cellular	PO_0009002
elaters	cellular	
female	morphological	PATO_0000383
gametophyte	morphological	PO_0028003
gemmae	cellular	PO_0025614
leaf apex	cellular	PO_0020137
leaf axis	cellular	
leaf base	cellular	PO_0020040
leaf center	cellular	
leaf cross section	cellular	
leaf lobes	cellular	
leaf margin	cellular	PO_0020128
leaf teeth	cellular	PO_0025518
leaf vitta	cellular	
male	morphological	PATO_0000384
papillae	cellular	NCIT_C113267
paraphyses	cellular	PO_0030070
perianth	anatomical	PO_0030031
perianth mouth	cellular	
perichaetial leaf	anatomical	PO_0030030
perigon leaf	anatomical	PO_0030028
plant apex	morphological	PO_0028004
postical lobe	anatomical	
rounded margin	cellular	
spores	cellular	PO_0025017
sporophyte	morphological	BTO_0000735
stature	morphological	
stature lateral side	morphological	
stature ventral side	morphological	
stem cross section	cellular	
toothed margin	cellular	PO_0025518

Table 1. List of phenotypic characters that were associated with the images. These biological terms allow to link ecologically relevant information to the images and *vice versa*. We ascertain that there is pressing need to extend controlled vocabularies such as the cell or plant ontologies with semantic terms specifically addressing phenotypic characters in bryophytes. This is because many existing terms were originally defined for vascular plants which have, for instance, dominant sporophytes and multicellular leaves which are different from the dominant gametophytes and unicellular leaves usually present in bryophytes.

two cogged wheels were coupled with a toothed belt to obtain very fine step increments of the stepping motor for high magnifications. A Cognisys StackShot controller was used to control the amount and distance of the stepping motor with the following controller settings: Dist/Rev: 3200 stp, Backlash: 0 steps, # pics: 1, Tsettle: 100.0 ms, Toff: 450.0 ms, Auto Return: yes, Speed: 3000 st/sec, Tlapse: off, Tpulse: 800.0 ms, Tramp: 100 ms, Units: steps, Torque: 6, Hi Precision: Off, LCD Backlight: 10, Mode: Auto-Step using between 25 steps (magnification 1x) and 50 steps (magnification 25x) and 100 steps (magnification 400x) (number of steps depending on aperture settings and effective magnification).

Raw images were recorded in CR3-format and pre-processed with Adobe Camera RAW. Non-destructive image processing such as corrections of the field curvature, removal of chromatic aberration, increase of contrast and brightness were performed in Adobe Camera RAW. Images were then exported to TIFF-format and any image processing steps were recorded in individual Adobe XMP-files.

Multi-focus image fusion was performed on the individual images in the z-stacks using the software Helicon Focus 7.7.5 and by choosing the algorithms *depth map* and *pyramid* with different settings of radius (4, 8, 16, 24) and smoothing (2, 4). The best composite images were chosen manually and retained. When composite images contained specimens that were larger than the frame, several images were stitched together using the *panorama stitching* function in the software Affinity Photo 1.10.1.

Semantic identifier	Name
{http://edamontology.org/data_1060}	File base name
{http://rs.tdwg.org/dwc/terms/acceptedScientificName}	Species name
{http://rs.tdwg.org/dwc/terms/TaxonID}	TaxonID
{http://rs.tdwg.org/dwc/terms/measurementMethod}	Measurement Method
{http://www.openmicroscopy.org/rdf/2016-06/ome_core/Instrument}	Microscope
{http://www.openmicroscopy.org/rdf/2016-06/ome_core/nominalMagnification}	Magnification
{http://www.openmicroscopy.org/rdf/2016-06/ome_core/ContrastMethod}	Contrast
{http://www.openmicroscopy.org/rdf/2016-06/ome_core/Objective}	Microscope Objective
{http://rs.tdwg.org/dwc/terms/basisOfRecord}	Basis of Record
{http://rs.tdwg.org/dwc/terms/PreservedSpecimen}	Voucher specimen
{http://rs.tdwg.org/dwc/terms/EarliestDateCollected}	Collection Date
{http://rs.tdwg.org/dwc/terms/recordedBy}	Collector
{http://rs.tdwg.org/dwc/terms/identifiedBy}	Determined
{http://rs.tdwg.org/dwc/terms/geodeticDatum}	Geodetic datum
{http://www.ebi.ac.uk/efo/EFO_0005020}	Latitude
{http://www.ebi.ac.uk/efo/EFO_0005021}	Longitude
{http://rs.tdwg.org/dwc/terms/verbatimElevation}	Elevation
{http://rs.tdwg.org/dwc/terms/coordinatePrecision}	Precision
{http://purl.obolibrary.org/obo/OBI_0001048}	Camera

Table 2. List of semantic terms used to annotate the segmented images.

Image segmentation. Images were manually segmented and interfering background removed using the *flood select*, *brush selection* and *freehand selection* tools in the software Affinity Photo. A stage micrometre was photographed separately with any of the objectives and microscope combinations to determine the scale which was then calculated per pixel for each combination (File *scale_bar_distances.csv* in⁵⁸). Scale bars were put post-hoc onto the segmented images using the Python script *scale_bar.py*⁵⁸.

Handling of metadata. Metadata including species name, taxonomic rank information (NCBI-Taxon, GBIF and OTT taxonomy identifiers), voucher specimen id, image acquisition date, an object description including the name of the captured phenotypic character(s), the used objective, microscope, and magnification were associated with any raw image based on unique respective file names. Table 1 lists the ecologically relevant phenotypic characters that were associated with the images. Individual file names (variable *file list*), name within an image focus stack (variable *stack name*) and name within an image stitching stack (variable *stitch name*) were recorded additionally to facilitate subsequent automatized image processing in computational workflows. A Python script was created to put individual images as part of image stacks into directories (File *create_image_stacks.py* in⁵⁹). The Python script parses the *Label* tag in the XMP-files. Any metadata regarding image enhancement and non-destructive image processing were extracted from XMP-files using a simple Python script (File *xmp_stack_to_tsv.py* in⁵⁹). The metadata was saved in individual TSV-files and merged using a helper Python script (File *tsv_merge.py* in⁵⁹). Supplementary Table 2 lists all fields which were extracted from the XMP.

Data deposition. Raw camera and pre-processed imaging data in CR3 and TIFF format, respectively, were uploaded to BioStudies using the command line IBM Aspera software tool *ascp* version 3.8.1.161274 to ensure that data has been transmitted without errors. Sparse file check summing was enabled to ensure integrity of files during transfer (parameter *-k 2*). The raw bioimaging data is available under the BioStudies identifier [S-BIAD188](#).

Pre-processed images were converted to the Bio-Formats OME-TIFF format⁶⁰ by creating intermediate ZARR-pyramid tiles using the *bioformats2raw* converter version 0.4.0 and then using the *raw2ometiff* version 0.3.0 software tool to create the final pyramid images. Individual fully segmented and processed images were associated with standardised geolocation information to improve data reuse and to enable linking bioimaging data to ecological data repositories. Swiss Topo CH1903/LV03 coordinates were converted to WGS84 using *Swisstopo-WGS84-LV03*⁶¹. The processed images were further associated with the metadata information listed in Table 2 to enable machine-readability in IDR. A helper script was implemented in R to facilitate the generation of TSV tables for data upload to the Image Data Resource (IDR) repository (*_tsv_res_2_idr.r* in⁶²). Processed images and the metadata aggregated in a TSV table were uploaded to IDR using the software *Globus Connect Personal* 3.1.6. The dataset is available under the identifier [idr0134](#).

Data Records

Two separate data records were created to enable rapid use of the data in machine learning and biodiversity approaches.

(1) The camera raw images (Canon CR3-format), the pre-processed images (16-bit TIFF-format), and the contextual metadata were deposited to BioStudies under the identifier [S-BIAD188](#)⁶³. The data record consists of a total of 223'989 individual raw image files partitioned into 48 species. The entire data record has a total size of approx. 12 TB.

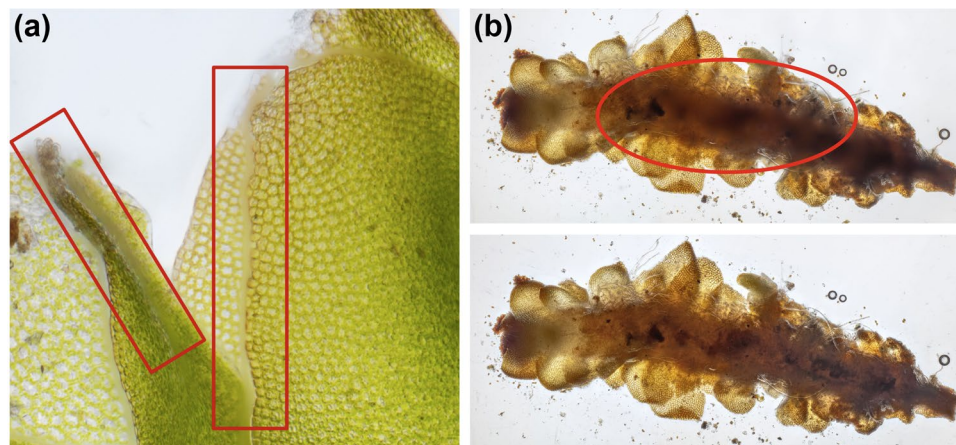


Fig. 4 Deficiencies of multi-focus image fusion methods. Red circles and bars were drawn *post-hoc* with Affinity Photo to indicate the deficient regions in the images, thus, the regions where multi-focus image fusion methods can produce blur and artefacts in composite images of microscopic objects. **(a)** Crop of *IMG_1532–1621 Scapania cuspiduligera stature ventral side*. Visible blur along edges (boundary regions) of overlapping leaves (Parameter settings: *Method: Depth Map, Radius: 8, Smoothing: 4*). **(b)** *IMG_0107–0226 Scapania ligulifolia stature dorsal side* (Parameter settings: *Method: Depth Map, Radius: 32, Smoothing: 20*).

(2) The pre-processed and fully segmented and processed images along with metadata were deposited in OME-TIFF format to the Image Data Resource (IDR) repository under the identifier [idr0134](https://doi.org/10.1038/s41597-022-01691-x)⁶⁴. The data record consists of a total of 4233 pre-processed and 905 fully processed imaged files. The data record has a total size of approx. 14 TB.

Technical Validation

Biological validation of species identity and visible phenotypic characters in the pre-processed images were performed by consulting the external experts Edwin Urmi, Heike Hofmann, Vadim Bakalin and Kristian Hassel. Photos of the herbarium specimen CM-30377 originating from North America (Supplementary Table 1) show quite different characters when compared to the voucher specimen B-108428 originating from Northern Europe. Hence, the taxonomic status of the species *Scapania glaucocephala* is not yet fully clear¹⁵. Photos of CM-30377 may, thus, relate to the species *Scapania scapanioides* (C.Massal.) Grolle, which is listed in⁷ as separate species occurring in Europe. Further, *S. brevicaulis* and *S. degenii* may comprise taxonomically identical species and additional research is needed to resolve their taxonomic status. Images from this study can help to clarify relationships of phenotypic characters and the phylogenetic and taxonomic status of cryptic species.

Multi-focused image fusion methods were applied with different settings to the individual images in stacks in order to validate the technical quality of fused composite images. Composite images were manually inspected and the best image retained. Generally, classic Laplacian pyramid transform-based methods such as *Pyramid Maximum Contrast* implemented in the software Helicon Focus produce good results in complex cases with regard to intersecting objects and along edges (boundary regions), but this algorithm increases contrast and glare and it is prone to noise and artefacts and is generally considered less accurate regarding the reproduction of microscopic objects^{65–68}. The deterministic depth map-based method implemented in the software Helicon Focus first calculates depth maps from intermediary images based on the absolute difference in the brightness of corresponding pixels in source images and smoothed intermediary images and then generates the composite image from the source image pixels with indices differing from the indices in the smoothed depth map⁶⁹. Whereas larger values for the parameter *radius* increase blur along edges, lower values can introduce artefacts, while the amount of blur along the transition between fused areas of individual images can be controlled with the parameter *smoothing*. The depth map-based method generally produces accurate reproductions of microscopic objects. However, in some circumstances and especially with high magnifications, it can generate large artefacts and blur around the edges (boundary regions) (Fig. 4). Recently, machine learning-based methods have been applied to focus-based image fusion tasks that may be superior to deterministic approaches³⁷. Although there have been proposed some algorithms specifically for microscopic imaging, there is a considerable lack of usable implementations and a lack of microscopic training data for machine learning-based algorithms³⁷. Our reference dataset can be used to train and improve these algorithms.

Python scripts have been written which are available as Open Source software in github^{58,59} to facilitate the automated processing of images. These scripts use metadata information to put individual images into image stacks to perform focus-based image fusion and image stitching tasks. However, most of the work has still been implemented manually and scientific workflows need to be developed that allow to fully automate the entire process combining images with software tools utilising the machine-actionable information contained in the metadata^{43,49,70}. Using the procedures described in^{20,46,47} metadata used herein has been validated. Standardised vocabularies were used following the FAIR guiding principles¹. When improved algorithms have been developed, the entire pipeline can be re-run resulting in improved segmented images without any further intervention. This data reuse and the rich documentation in metadata will foster good scientific practices through source tracking and provenance.

Usage Notes

Raw camera and pre-processed imaging data, the fully segmented and processed images, and the metadata are available under the terms of the license Creative Commons CC BY 4.0. Open-Source software scripts and code^{58,59} are available under the terms of the BSD 3-Clause license.

Code availability

Software code and scripts used in this study are available as Open Source in github^{58,59,62}. Python scripts were tested under Python 3.7 and require the additional modules PIL, pandas, xml, csv, errno, sys, os, argparse, glob, pathlib and re. R scripts were tested in R 4.1.3 and require the additional packages parallel, foreach, and doMC. Shell scripts were tested using Bourne Again Shell (bash) 5.1.16.

Received: 17 May 2022; Accepted: 8 September 2022;

Published online: 04 October 2022

References

1. Wilkinson, M. D. *et al.* The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data* **3**, 160018 (2016).
2. Ellenberg, J. *et al.* A call for public archives for biological image data. *Nat Methods* **15**, 849–854 (2018).
3. Löffler, F., Wesp, V., König-Ries, B. & Klan, F. Dataset search in biodiversity research: Do metadata in data repositories reflect scholarly information needs? *PLoS ONE* **16**, e0246099 (2021).
4. Asakawa, Y., Ludwiczuk, A. & Nagashima, F. Phytochemical and biological studies of bryophytes. *Phytochemistry* **91**, 52–80 (2013).
5. He, X., Sun, Y. & Zhu, R.-L. The Oil Bodies of Liverworts: Unique and Important Organelles in Land Plants. *Critical Reviews in Plant Sciences* **32**, 293–302 (2013).
6. Kanazawa, T. *et al.* The liverwort oil body is formed by redirection of the secretory pathway. *Nat Commun* **11**, 6152 (2020).
7. Hodgetts, N. G. *et al.* An annotated checklist of bryophytes of Europe, Macaronesia and Cyprus. *Journal of Bryology* **42**, 1–116 (2020).
8. Spitale, D. Switch between competition and facilitation within a seasonal scale at colony level in bryophytes. *Oecologia* **160**, 471–482 (2009).
9. Andersen, N. H. *et al.* Sesquiterpenes of nine European liverworts from the genera, *Anastrepta*, *Bazzania*, *Jungermannia*, *Lepidozia* and *Scapania*. *Phytochemistry* **16**, 1731–1751 (1977).
10. Guo, L. *et al.* Chemical Composition, Antifungal and Antitumor Properties of Ether Extracts of *Scapania verrucosa* Heeg. and its Endophytic Fungus *Chaetomium fusiforme*. *Molecules* **13**, 2114–2125 (2008).
11. Bukvicki, D. R. *et al.* Assessment of the Chemical Composition and *In Vitro* Antimicrobial Potential of Extracts of the Liverwort *Scapania Aspera*. *Natural Product Communications* **8**, 1934578X1300800 (2013).
12. Han, J. *et al.* Terpenoids from Chinese Liverworts *Scapania* spp. *J. Nat. Prod.* **84**, 1210–1215 (2021).
13. Vázquez, M. D., López, J. & Carballeira, A. Uptake of Heavy Metals to the Extracellular and Intracellular Compartments in Three Species of Aquatic Bryophyte. *Ecotoxicology and Environmental Safety* **44**, 12–24 (1999).
14. Samecka-Cymerman, A., Kolon, K. & Kempers, A. J. Heavy Metals in Aquatic Bryophytes from the Ore Mountains (Germany). *Ecotoxicology and Environmental Safety* **52**, 203–210 (2002).
15. Heinrichs, J. *et al.* A phylogeny of the northern temperate leafy liverwort genus *Scapania* (Scapaniaceae, Jungermanniales). *Molecular Phylogenetics and Evolution* **62**, 973–985 (2012).
16. Vana, J., Hentschel, J., Müller, J. & Heinrichs, J. Taxonomic novelties in *Scapania*. *PhytoKeys* **10**, 13 (2012).
17. Choi, S. S., Min, J., Kwon, W. & Park, J. The complete mitochondrial genome of *Scapania ampliata* Steph., 1897 (Scapaniaceae, Jungermanniales). *Mitochondrial DNA B Resour.* **6**, 686–688 (2021).
18. Choi, S. S., Bakalin, V. A., Kwon, W. & Park, J. The complete mitochondrial genome of *Douinia plicata* (Lindb.) Konstant. *et. Vilnet* (Scapaniaceae, Jungermanniales). *Mitochondrial DNA B Resour.* **6**, 789–791 (2021).
19. Bernhardt-Römermann, M., Poschold, P. & Hentschel, J. BryForTrait - A life-history trait database of forest bryophytes. *J Veg Sci* **29**, 798–800 (2018).
20. Schneider, F. D. *et al.* Towards an ecological trait-data standard. *Methods Ecol Evol* **10**, 2006–2019 (2019).
21. Kommineni, V. K. *et al.* Comprehensive leaf size traits dataset for seven plant species from digitised herbarium specimen images covering more than two centuries. *BDJ* **9**, e69806 (2021).
22. Meijering, E., Carpenter, A. E., Peng, H., Hamprecht, F. A. & Olivo-Marin, J.-C. Imagining the future of bioimage analysis. *Nat Biotechnol* **34**, 1250–1255 (2016).
23. Cornelissen, J. H. C., Lang, S. L., Soudzilovskaia, N. A. & During, H. J. Comparative Cryptogam Ecology: A Review of Bryophyte and Lichen Traits that Drive Biogeochemistry. *Annals of Botany* **99**, 987–1001 (2007).
24. Diaz, S. *et al.* The global spectrum of plant form and function. *Nature* **529**, 167–171 (2016).
25. Brodribb, T. J., Carriqui, M., Delzon, S., McAdam, S. A. M. & Holbrook, N. M. Advanced vascular function discovered in a widespread moss. *Nat. Plants* **6**, 273–279 (2020).
26. Duckett, J. G. & Pressel, S. Of mosses and vascular plants. *Nat. Plants* **6**, 184–185 (2020).
27. Horn, A. *et al.* Natural Products from Bryophytes: From Basic Biology to Biotechnological Applications. 28.
28. Schindel, D. E. & Cook, J. A. The next generation of natural history collections. *PLoS Biol* **16**, e2006125 (2018).
29. Hedrick, B. P. *et al.* Digitization and the Future of Natural History Collections. *BioScience* **70**, 243–251 (2020).
30. Gurevitch, J., Koricheva, J., Nakagawa, S. & Stewart, G. Meta-analysis and the science of research synthesis. *Nature* **555**, 175–182 (2018).
31. Peters, K., Gorzalka, K., Bruehlheide, H. & Neumann, S. Seasonal variation of secondary metabolites in nine different bryophytes. *Ecology and Evolution* **8**, 9105–9117 (2018).
32. Valdecasas, A. G., Marshall, D., Becerra, J. M. & Terrero, J. J. On the extended depth of focus algorithms for brightfield microscopy. *Micron* **32**, 559–569 (2001).
33. Goldsmith, N. T. Deep Focus; A Digital Image Processing Technique To Produce Improved Focal Depth In Light Microscopy. *Image Anal Stereol* **19**, 163 (2011).
34. Nasibov, A., Nasibov, H. & Hacizade, F. Seamless image stitching algorithm using radiometric lens calibration for high resolution optical microscopy. In *2009 Fifth International Conference on Soft Computing, Computing with Words and Perceptions in System Analysis, Decision and Control* 1–4, <https://doi.org/10.1109/ICSCCW.2009.5379500> (IEEE, 2009).
35. Wang, Z. & Yang, Z. Review on image-stitching techniques. *Multimedia Systems* **26**, 413–430 (2020).
36. Piper, J. Software-Based Stacking Techniques to Enhance Depth of Field and Dynamic Range in Digital Photomicrography. in *Histology Protocols* (eds. Hewitson, T. D. & Darby, I. A.) vol. 611 193–210 (Humana Press, 2010).
37. Liu, Y., Wang, L., Cheng, J., Li, C. & Chen, X. Multi-focus image fusion: A Survey of the state of the art. *Information Fusion* **64**, 71–91 (2020).
38. Yang, J., Wright, J., Huang, T. S. & Ma, Y. Image Super-Resolution Via Sparse Representation. *IEEE Trans. on Image Process.* **19**, 2861–2873 (2010).
39. Yin, H., Li, S. & Fang, L. Simultaneous image fusion and super-resolution using sparse representation. *Information Fusion* **14**, 229–240 (2013).

40. Yu, Z., Liu, S., Zhu, D., Kuang, C. & Liu, X. Parallel detecting super-resolution microscopy using correlation based image restoration. *Optics Communications* **404**, 139–146 (2017).
41. Yang, B., Zhong, J., Li, Y. & Chen, Z. Multi-focus image fusion and super-resolution with convolutional neural network. *Int. J. Wavelets Multiresolut Inf. Process.* **15**, 1750037 (2017).
42. Peters *et al.* PhenoMeNal: processing and analysis of metabolomics data in the cloud. *GigaScience* **8** (2019).
43. Goble, C. *et al.* FAIR Computational Workflows. *Data Intelligence* **2**, 108–121 (2020).
44. Atkinson, M., Gesing, S., Montagnat, J. & Taylor, I. Scientific workflows: Past, present and future. *Future Generation Computer Systems* **75**, 216–227 (2017).
45. Miksa, T., Simms, S., Mietchen, D. & Jones, S. Ten principles for machine-actionable data management plans. *PLoS Comput Biol* **15**, e1006750 (2019).
46. Samuel, S., Taubert, F., Walther, D., König-Ries, B. & Bücker, H. M. Towards Reproducibility of Microscopy Experiments. *D-Lib Magazine* **23** (2017).
47. Kunis, S. *et al.* MDEmic: a metadata annotation tool to facilitate management of FAIR image data in the bioimaging community. *Nat Methods*, <https://doi.org/10.1038/s41592-021-01288-z> (2021).
48. Samuel, S. & König-Ries, B. End-to-End provenance representation for the understandability and reproducibility of scientific experiments using a semantic approach. *J Biomed Semant* **13**, 1 (2022).
49. Wratten, L., Wilm, A. & Göke, J. Reproducible, scalable, and shareable analysis pipelines with bioinformatics workflow managers. *Nat Methods* **18**, 1161–1168 (2021).
50. Hansen, O. L. P. *et al.* Species-level image classification with convolutional neural network enables insect identification from habitus images. *Ecol Evol* **10**, 737–747 (2020).
51. Høye, T. T. *et al.* Deep learning and computer vision will transform entomology. *Proc Natl Acad Sci USA* **118**, e2002545117 (2021).
52. König, C. *et al.* Biodiversity data integration—the significance of data resolution and domain. *PLoS Biol* **17**, e3000183 (2019).
53. Nelson, G. & Paul, D. L. DiSSCo, iDigBio and the Future of Global Collaboration. *BISS* **3**, e37896 (2019).
54. Culina, A. *et al.* Navigating the unfolding open data landscape in ecology and evolution. *Nat Ecol Evol* **2**, 420–426 (2018).
55. Seltzer, C. Making Biodiversity Data Social, Shareable, and Scalable: Reflections on iNaturalist & citizen science. *BISS* **3**, e46670 (2019).
56. Borgman, C. L. & Bourne, P. E. Why it takes a village to manage and share data. *Harvard Data Science Review* **4**(3) (2022).
57. Holmgren, P. K. & Holmgren, N. H. Index Herbariorum. *Taxon* **40**, 687–692 (1991).
58. Peters, K. Draw a scale bar on microscopic images. *Zenodo*. <https://doi.org/10.5281/ZENODO.5592446> (2021).
59. Peters, K. Create script to build image stacks based on a list of XMP files containing color badges. *Zenodo*. <https://doi.org/10.5281/ZENODO.5592436> (2021).
60. Besson, S. *et al.* Bringing Open Data to Whole Slide Imaging. in *Digital Pathology* (eds. Reyes-Aldasoro, C. C., Janowczyk, A., Veta, M., Bankhead, P. & Sirinukunwattana, K.) vol. 11435 3–10 (Springer International Publishing, 2019).
61. Marti, U. & Dupraz, H. Swisstopo Scripts GPS WGS84 <-> LV03 (CH1903). (2021).
62. Peters, K. Scripts for bioimage submission. *Zenodo*. <https://doi.org/10.5281/ZENODO.6447017> (2022).
63. Peters, K. Reference raw BioImaging dataset to assess the phenotypic trait diversity of bryophytes within the family Scapaniaceae. *BioStudies*. <https://www.ebi.ac.uk/biostudies/studies/S-BIAD188> (2022).
64. Peters, K. Reference BioImaging dataset to assess the phenotypic trait diversity of bryophytes within the family Scapaniaceae. *Image Data Resource (University of Dundee)*. <https://doi.org/10.17867/10000183> (2022).
65. Adelson, E. H., Anderson, C. H., Bergen, J. R., Burt, P. J. & Ogden, J. M. Pyramid methods in image processing. *RCA Engineer* **29**, 33–41 (1984).
66. Ogden, J. M., Adelson, E. H., Bergen, J. R. & Burt, P. J. Pyramid-based computer graphics. *RCA Engineer* **30**, 4–15 (1985).
67. Toet, A. Image fusion by a ratio of low-pass pyramid. *Pattern Recognition Letters* **9**, 245–253 (1989).
68. Liu, Z., Tsukada, K., Hanasaki, K., Ho, Y. K. & Dai, Y. P. Image fusion by using steerable pyramid. *Pattern Recognition Letters* **22**, 929–939 (2001).
69. Kozub, D. Focus stacking of captured images. US Patent. 10,389,936 B2 (2019).
70. Perkel, J. M. Workflow systems turn raw data into scientific knowledge. *Nature* **573**, 149–150 (2019).
71. Urmi, E., Hofmann, H. & Schubiger, C. *Scapania aspera* Bernet & M.Bernet. <https://doi.org/10.5167/UZH-197490> (2020).
72. Urmi, E., Hofmann, H. & Schubiger, C. *Scapania subalpina* (Lindenb.) Dumort. <https://doi.org/10.5167/UZH-197517> (2020).
73. Urmi, E., Hofmann, H. & Schubiger, C. *Scapania undulata* (L.) Dumort. <https://doi.org/10.5167/UZH-197522> (2020).
74. Urmi, E., Peters, K. & Schubiger, C. *Scapania nemorea* subsp. *nemorea* (L.) Grolle. <https://doi.org/10.5167/UZH-205668> (2020).
75. Urmi, E., Peters, K. & Schubiger, C. *Scapania aequiloba* (Schwägr.) Dumort. <https://doi.org/10.5167/UZH-197488> (2020).
76. Urmi, E., Peters, K. & Schubiger, C. *Scapania apiculata* Spruce. <https://doi.org/10.5167/UZH-197489> (2020).
77. Urmi, E., Peters, K. & Schubiger, C. *Scapania calcicola* (Arnell & J.Perss.) Ingham. <https://doi.org/10.5167/UZH-197492> (2020).
78. Urmi, E., Peters, K. & Schubiger, C. *Scapania carinthiaca* Lindb. <https://doi.org/10.5167/UZH-197494> (2020).
79. Urmi, E., Peters, K. & Schubiger, C. *Scapania compacta* (Roth) Dumort. <https://doi.org/10.5167/UZH-197496> (2020).
80. Urmi, E., Peters, K. & Schubiger, C. *Scapania curta* (Mart.) Dumort. <https://doi.org/10.5167/UZH-197497> (2020).
81. Urmi, E., Peters, K. & Schubiger, C. *Scapania cuspiduligera* (Nees) Müll.Frib. <https://doi.org/10.5167/UZH-197499> (2020).
82. Urmi, E., Peters, K. & Schubiger, C. *Scapania degenii* Müll.Frib. <https://doi.org/10.5167/UZH-197500> (2020).
83. Urmi, E., Peters, K. & Schubiger, C. *Scapania gracilis* Lindb. <https://doi.org/10.5167/UZH-197503> (2020).
84. Urmi, E., Peters, K. & Schubiger, C. *Scapania gymnostomophila* Kaal. <https://doi.org/10.5167/UZH-197504> (2020).
85. Urmi, E., Peters, K. & Schubiger, C. *Scapania helvetica* Gottsche. <https://doi.org/10.5167/UZH-197505> (2020).
86. Urmi, E., Peters, K. & Schubiger, C. *Scapania irrigua* subsp. *irrigua* (Nees) Nees. <https://doi.org/10.5167/UZH-197506> (2020).
87. Urmi, E., Peters, K. & Schubiger, C. *Scapania irrigua* subsp. *rufescens* (Loeske) R.M.Schust. <https://doi.org/10.5167/UZH-197507> (2020).
88. Urmi, E., Peters, K. & Schubiger, C. *Scapania mucronata* subsp. *mucronata* H.Buch. <https://doi.org/10.5167/UZH-197508> (2020).
89. Urmi, E., Peters, K. & Schubiger, C. *Scapania mucronata* subsp. *praetervisa* (Meyl.) R.M.Schust. <https://doi.org/10.5167/UZH-197509> (2020).
90. Urmi, E., Peters, K. & Schubiger, C. *Scapania obscura* (Arnell & C.E.O.Jensen) Schiffn. <https://doi.org/10.5167/UZH-197511> (2020).
91. Urmi, E., Peters, K. & Schubiger, C. *Scapania paludicola* Loeske & Müll.Frib. <https://doi.org/10.5167/UZH-197513> (2020).
92. Urmi, E., Peters, K. & Schubiger, C. *Scapania paludosa* (Müll.Frib.) Müll.Frib. <https://doi.org/10.5167/UZH-197514> (2020).
93. Urmi, E., Peters, K. & Schubiger, C. *Scapania scandica* (Arnell & H.Buch) Macvicar. <https://doi.org/10.5167/UZH-197515> (2020).
94. Urmi, E., Peters, K. & Schubiger, C. *Scapania uliginosa* (Lindenb.) Dumort. <https://doi.org/10.5167/UZH-197520> (2020).
95. Urmi, E., Peters, K. & Schubiger, C. *Scapania umbrosa* (Schrad.) Dumort. <https://doi.org/10.5167/UZH-197521> (2020).
96. Urmi, E., Peters, K. & Schubiger, C. *Scapania verrucosa* Heeg. <https://doi.org/10.5167/UZH-197523> (2020).

Acknowledgements

KP acknowledges the support of iDiv (funded by the German Research Foundation, DFG-FZT 118, 202548816) and Swissbryophytes. Fully segmented and processed images of 27 *Scapania* species occurring in Switzerland were originally produced for the website <http://www.swissbryophytes.ch> and are also available at the Zurich Open Repository Archive^{71–96}. We would also like to thank the external experts Edwin Urmi, Heike Hofmann, Vadim Bakalin, Kristian Hassel (Bryophyte herbarium at TRH, NTNU University Museum), and the Bryophyte Herbarium at the University of

Copenhagen for providing voucher specimens and for performing validation of species identities on images of visible traits and especially Edwin Urmi for help with identification of fresh samples.

Author contributions

K.P. performed the entire study. B.K.R. supervised the study.

Funding

Open Access funding enabled and organized by Projekt DEAL.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41597-022-01691-x>.

Correspondence and requests for materials should be addressed to K.P.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2022