Scientific Data

https://doi.org/10.1038/s41597-026-07064-y

Article in Press

From paper to pixels - digitized maps of vascular plant distributions in Denmark in the early 20th century

Received: 17 November 2025

Accepted: 11 March 2026

Cite this article as: Christiansen, D.M., Arnell, M., Lange, C.B. *et al.* From paper to pixels - digitized maps of vascular plant distributions in Denmark in the early 20th century. *Sci Data* (2026). https://doi.org/10.1038/s41597-026-07064-y

Ditte Marie Christiansen, Matilda Arnell, Conny Bruun Asmussen Lange, Tankred Ott, Hans Henrik Bruun & Sergey Rosbakh

From paper to pixels - digitized maps of vascular plant distributions in Denmark in the early 20th century

Ditte Marie Christiansen[1,*], Matilda Arnell[2], Conny Bruun Asmussen Lange[1], Tankred Ott[3], Hans Henrik Bruun[4,+] and Sergey Rosbakh[1,+]

Affiliations

1. Department of Plant and Environmental Sciences, University of Copenhagen, Frederiksberg C, Denmark
2. Department of Ecology, Environment and Plant Sciences, Stockholm University, Stockholm, Sweden
3. Novo Nordisk A/S, Bagsværd, Denmark.
4. Department of Biology, University of Copenhagen, Copenhagen Ø, Denmark

*corresponding author(s): Ditte Marie Christiansen (dmc@plen.ku.dk/dittechristiansen@live.dk)

+ Hans Henrik Bruun and Sergey Rosbakh are equally contributing senior authors

Abstract

Historic data are essential for quantifying changes in biodiversity and their drivers. Much legacy data remains undigitized in archives, such as natural history museums, representing untapped resources for assessing past biodiversity patterns. In 1904, the first systematic survey of species distributions of the entire Danish vascular flora was initiated. Field surveys were essentially conducted in the early decades of the 20th century. Results were published as black and white printed maps (1931-1976). We used automated georeferencing and image classification to extract data from all published maps. We present 1340 historic distribution maps of vascular plant species, subspecies and varieties in Denmark at 10 × 10 km resolution. These maps represent the Danish flora of native and archaeophyte species known at the time, except a dozen ubiquitous native species. By comparison with modern distribution data, long-term increases and declines can be accurately quantified, contributing to large-scale analysis of national, European or global plant diversity changes in the modern era.

Background & Summary

Human use and alteration of natural ecosystems and organisms have increased rapidly in the last century[1]. Pressures such as habitat destruction, eutrophication, and climate change have further led to population declines and local extirpation of species, resulting in increased threat levels of species globally[2]. Biodiversity is changing worldwide, but documenting changes consistently across scales has proven difficult[3]. For many regions, biodiversity change may be severely underestimated, as most surveillance data extend only a few decades back. Even in the best cases, systematic monitoring began long after the onset of large-scale industrialization of agriculture and forestry[4], which constitutes the beginning of the third phase of the Anthropocene[5]. Using data from only a few decades ago, many studies might therefore have fallen short of accurately assessing biodiversity change and potentially underestimated biodiversity trends and anthropogenic impacts (i.e. enforcing shifting baseline syndrome[6]).

Knowledge of occupancy of plant species before the mid-20th century upsurge in land-use intensification is necessary to estimate the extent and direction of biodiversity changes in the recent past[7]. Resurveys of historical vegetation plots have been used to document changes in species elevational limits[8,9], plant community composition[10], and to examine large scale direction and extent of biodiversity change[11–13]. However, data on region-wide occupancy and on longer time scales are still in very short supply. A large quantity of historic data on species occurrences is stored undigitized in museums and naturalists' societies as physical publications, unpublished dot maps, herbarium specimens, and floristics reports, making such archives untapped resources that have the potential to fill this important data gap. For Denmark, one such data source, and arguably the most comprehensive, is a collection of publications of species treatises and occurrence maps resulting from a national survey of vascular plant species dating back to the beginning of the 1900s.

Topographical-Botanical Investigation

In 1904, the Danish Botanical Society initiated a nationwide mapping of the Danish flora, i.e., the Topographical-Botanical Investigation (Topografisk-Botanisk Undersøgelse; TBU)[14]. The investigation was organized as county-wise surveys, with 47 approximately equal-sized districts largely based on deaneries (clerical counties) and often delimited by natural borders such as rivers. The 47 original districts were amended with six more by the 1920 re-unification of Northern Schleswig with Denmark. At the same time, each of the four largest original districts were split in two, making a total of 57 districts. More than 100 society members and local amateur botanists, typically school teachers, parish ministers, pharmacists and doctors, assumed responsibility for their home region[14,15]. From 1904 to 1923, the contributors submitted reports containing enumerations of county floras, including named occurrence sites of species and, for common species, an estimated local occupancy, following guidelines issued by the organization committee. When in doubt, the contributors also sent in specimens for further scrutiny, which greatly increased the Botanical Museum's herbarium collection (now Natural History Museum of Denmark)[14]. The curator of the Danish Herbarium at the University of Copenhagen transferred the information from all existing herbarium sheet labels to index cards, one card per species per district, to which also records from published local floras, excursion reports, records reported in letters to the curator and records from the county surveys were added[14]. In 1931, the first TBU report was published, describing and mapping the distribution of all Danish species of Fabaceae[16]. The background and methodology were laid out same year by C. H. Ostenfeld[14]. These publications were in English and by university professors. All subsequent reports, issued family by family, were in Danish and many were written by non-university botanists, such as the teacher's college reader and trained botanist Anfred Pedersen (1920-2004). The subsequent Danish Herbarium curators continued to amass records from excursion reports, letters and newly submitted herbarium specimens, enlarging the knowledge base for the later issued reports. The completion of the series took considerably longer than anticipated. The final report, on Portulacaceae and Valerianaceae was issued in 1976, 45 years after the first publication, but still in large part based on the surveys undertaken 1904-1925. A report mimicking the TBU format was published in 1980, mapping *Rubus* microspecies from herbarium material and undertaken in essence by Anfred Pedersen[15]. In total, the series including the report on Rubus microspecies comprises distribution maps of 1401 vascular plant species, subspecies and varieties, published in family-wise treatises from 1931-1980 (Table 1). To increase availability to an international audience, a selection of the maps was compiled and reprinted in facsimile by Vestergaard and Hansen (1989)[17], accompanied by introductory chapters on climate, dominant soil types and other environmental properties.

Table 1. Overview of all Topographical-Botanical Survey (TBU) publications. The survey results with maps were published by plant family from 1931 to 1976. A publication mimicking the TBU publications containing microspecies of *Rubus* family were published in 1980. Plant family is noted in the table as given in the TBU publications together with author name and year of the publication.

Plant family	**Author**	**Year**
Papilionaceae	K. Jessen	1931[16]

Amaryllidaceae, Colchicaceae, Convallariaceae, Iridaceae, Liliaceae	K. Jessen	1935[18]
Primulaceae	T. Sørensen	1935[19]
Empetraceae, Ericaceae, Vacciniaceae	T. W. Böcher	1937[20]
Juncaceae	K. Wiinstedt	1937[21]
Polygonaceae	J. Grøntved	1939[22]
Hypericaceae	Aa. Køie	1939[23]
Araceae, Droseraceae, Geraniaceae, Lemnaceae	Aa. Køie & M. Køie	1939[24]
Compositae (Hieracium)	K. Wiindstedt	1939[25]
Cyperaceae (Scirpoideae)	K. Wiindstedt	1943[26]
Alismatacea, Juncaginaceae,Hydrocharitaceae	V. M. Mikkelsen	1943[27]
Crassulaceae, Saxifragaceae	M. Hoff	1943[28]
Plumbaginaceae, Pyrolaceae	A. Andersen	1943[29]
Cyperaceae (Caricoideae)	K. Wiinstedt	1945[30]
Campanulaceae, Lobeliaceae	A. Hansen	1948[31]
Orchidaceae	J. Grøntved	1948[32]
Umbelliferae	B. Egeholm	1951[33]
Adoxaceae, Caprifoliaceae, Cucurbitaceae, Dipsacaceae	A. Hansen	1951[34]
Equisetaceae, Isoëtaceae, Lycopodiaceae, Marsiliaceae, Ophioglossaceaè, Osmundaceae, Polypodiaceae, Selaginellaceae	K. Wiinstedt	1953[35]
Euphorbiaceae, Malvaceae, Violaceae	S. M. Rasmussen	1954[36]
Sparganiaceae, Typhaceae	J. Grøntved	1954[37]
Balsaminaceae, Linaceae, Oxalidaceae, Polygalaceae, Rubiaceae	A. Pedersen	1956[38]
Ranunculaceae	K. Larsen	1956[39]
Cruciferae	A. Pedersen	1958[40]
Asclepiadaceae, Apocynaceae, Gentianaceae, Menyanthaceae	A. Hansen	1958[41]
Caryophyllaceae	A. Pedersen	1959[42]
Lentibulariaceae, Plantaginaceae	A. Hansen	1960[43]
Ceratophyllaceae, Elatinaceae, Fumariaceae, Halorrhagidaceae, Hippuridaceae, Lythraceae, Nymphaeaceae, Papaveraceae	K. Larsen & A. Pedersen	1960[44]
Compositae (excl. Hieracium og Taraxacum)	A. Pedersen	1961[45]
Orobanchaceae, Scrophulariaceae	A. Pedersen	1963[46]
Convolvulaceae, Cuscutaceae, Hydrophyllaceae, Polemoniaceae, Solanaceae	A. Hansen	1963[47]
Boraginaceae	S. M. Rasmussen	1965[48]
Rosaceae (excl. Rubus p.p.)	A. Pedersen	1965[49]
Aristolochiaceae, Cannabaceae, Cistaceae, Resedaceae, Santalaceae, Urticaceae	A. Pedersen	1966[50]
Callitrichaceae	A. Pedersen	1966[51]
Amaranthaceae, Chenopodiaceae	A. Hansen & A. Pedersen	1968[52]

Aceraceae, Aquifoliacea, Araliacea, Berberidaceae, Betulacea, Celastracea, Cornaceae, Corylacea, Cupressaceae, Elaeagnacea, Fagaceae, Grossulariacea, Hippocastanacea, Loranthaceae, Myricaceae, Oleaceae, Pinaceae, Rhamnacea, Rosaceae, Salicacea, Taxaceae, Thymelaeaceae, Tiliaceae, Ulmaceae	S. Ødum	1968[53]
Labiatae, Verbenaceae	A. Pedersen	1969[54]
Oenotheraceae	N. Jensen	1971[55]
Gramineae (native and naturalized species)	A. Pedersen	1974[56]
Najadaceae, Potamogetonaceae, Ruppiaceae, Zannichelliaceae, Zosteraceae	A. Pedersen	1976[57]
Portulacaceae, Valerianaceae	A. Hansen & A. Pedersen	1976[58]
Rosaceae (Rubus)	A. Pedersen	1980[59]

The TBU publications are all public domain but the maps have up until now not been georeferenced or digitized, not been made broadly available, and not been used in any quantitative analysis of change. Some maps have been used in the Danish Red List assessments to assess occupancy change by manually counting occurrence points[60] and for a visual comparison with a modern survey of species occupancy[61]. To close this gap, we used automated georeferencing and image classification to convert the TBU maps into spatially explicit information on historic occupancy patterns. The plant occurrence data were assigned to a 10 × 10 km grid net. The maps cover approximately 1200 recognised Danish vascular plant species, representing the Danish flora of native species and archaeophytes known at the time of the TBU survey, with a few exceptions amongst the most common and widespread native species[62]. We provide the methodology used for digitizing the spatial distribution of plant species from historic maps and a thorough taxon list with synonyms accompanying the maps, making it possible to compare these historic plant species distribution data to modern species distribution data.

Methods

Map symbols

Each TBU map has the same base map of the coast line and land borders of Denmark surrounded by a black frame with coordinates. All maps are black and white and include a framed insertion of the island of Bornholm. Each map includes a small set of standardized symbols representing different levels of evidence for presence (or absence). The most common symbols are presented in Figure 1 and Table 2 along with the presence or absence category. The first presence category includes symbols that represent reliable presences of the species as actual herbarium specimens or flora lists deemed reliable. Expert-estimated abundance of “common to very common” is also assumed to represent a reliable species presence in that area. The second presence category is an expert-estimated regional occurrence as “here and there”. The first category of absence is made up of places where the focal species has been found in the past, but has since disappeared. The second category of absence is grid cells with no symbols or hatching, representing complete absence of the species (Table 2).

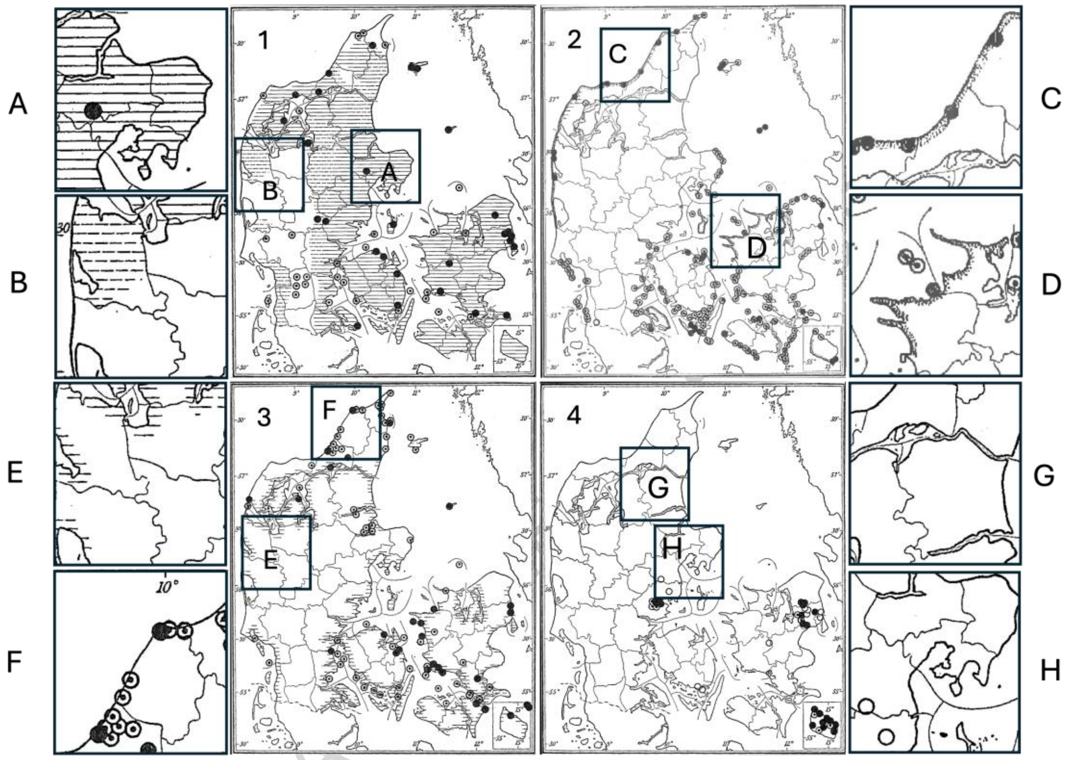


Figure 1. Examples of four historic maps from the Topographical-Botanical Survey (TBU) for 1) *Lysimachia arvensis*, 2) *Eryngium maritimum,* 3) *Lysimachia maritima*, and 4) *Primula farinosa*. Black dots and ‘bird’s eye’ i.e. open circles with a dot (Figure 1A, C, D, F) represent records based on herbarium specimens and records based on TBU reports, letters or published literature, respectively. Full hatching (Figure 1A), with or without dot symbols, indicates presence and expert-estimated abundance “common to very common”. Dashed hatching (Figure 1B), without dot symbols, indicates presence and expert-estimated abundance “here and there”. No hatching in areas with scattered dot symbols may be taken to indicate presence as “rare or rather rare.” As shown in Figure 1C, 1D and 1E, presence of species restricted to coastal habitats can be depicted as hatching or fine dotting along the coastline. All three are categorised as present (category 1). Figure 1G shows the background map without any occurrence information, categorised as absence of the focal species by our digitization procedure. In some published maps, in particular for nationally rare or noteworthy species, open circles (Figure 1H) show an older (before the TBU) record of occurrence, from which the species has later disappeared (well before the year of publication).

Table 2. Descriptions of the standardised symbols used for presence and absence of plant species in the maps from the Topographical-Botanical Survey (TBU).

Symbol	Category	Definition
Black dots (Fig. 1A, C, D, F)	Presence 1	A record based on a herbarium specimen. A dot can only count in one 10 × 10 km grid cell
Bird's eye dots, i.e. open circles with a centre point (Fig 1D and F)	Presence 1	A record based on the TBU district survey reports, letters or published literature. A dot can only count in one 10 × 10 km grid cell
Full black lines (Fig. 1A)	Presence 1	Expert-estimated abundance "common to very common"
Full black lines or hatching along a coastline (Fig. 1E)	Presence 1	Continuous presence of coastal species
Fine black dotting along the coastline (Fig. 1C and D)	Presence 1	Presence of coastal species
Dashed hatching (Fig. 1B)	Presence 2	Expert-estimated abundance "here and there"
Open dots (Fig. 1H)	Absence 1	The species has been found there in the past, but has disappeared
Background image (Fig. 1G)	Absence 2	Grid cells with no presence symbol was annotated as absence

Digitizing and georeferencing

We obtained the physical original published maps and pdf-scans of the maps from the archive of the Danish Botanical Society, in which copies of all TBU publication are stored. Some pdf-scans were of too low quality for vectorization, in which case we scanned the physical publication to pdfs in a RICOH IM C5510A scanner with 600 dpi resolution. We converted the scanned pdfs to png format keeping the resolution in the conversion. In the published treatises, one to four map panels are depicted on each page. Each map has the same baseline map of Denmark with a black frame (Fig. 1). We wrote a macro code for the program Fiji[63] to automatically detect one to four map frames per image, crop the image to the border of the frame, and save a separate image file (jpeg format) for each map (Fig. 2). The first step of scanning the physical maps sometimes resulted in slight differences in angles and sizes of the maps. To automate the later georeferencing of the maps (Fig. 2), we wrote another macro for the program Fiji that automatically aligned the maps in angle and size using a reference background map (for both macros, see [64]). We manually georeferenced the reference map using QGIS v. 3.32.3-Lima[65] and placed 33 points to cover most directions of the coastline of Denmark. Based on the georeferenced sites on the reference map, we automated the georeferencing of the other maps in R version 4.4.1[66] using the packages *gtools*[67] and *gdalUtilities*[68]. All georeferenced maps were saved in tiff format and are made available at the Figshare repository[69].

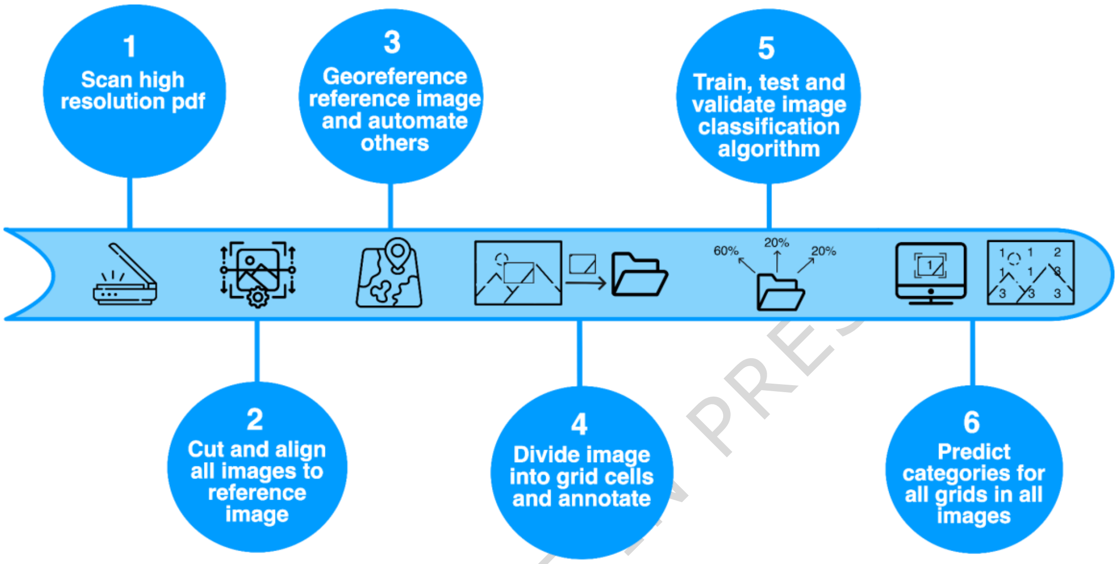


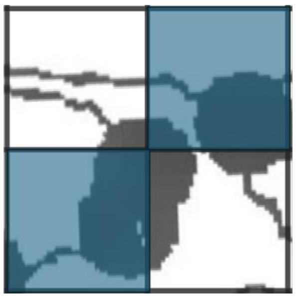
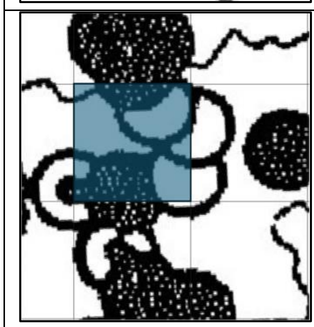
Figure 2. Conceptual figure of the TBU maps digitization process. 1) Scanning of the map pages to high resolution (600 dpi) pdf format. We converted the pdf map pages to png format keeping the resolution to 600 dpi. 2) Maps are extracted from the page and are aligned in angle and size based on a reference map using written macros in Fiji. From Fiji, all map images are saved as jpeg files. 3) Georeferencing of the reference map with 33 points located all around the Danish coastline. The georeferenced map coordinates are used to automatise the georeferencing for all other maps. 4) Division of georeferenced maps into 10 × 10 km grid cells and annotation of each grid cell image of 50 random maps after distribution category (species absence, species presence 1 and species presence 2). 5) Annotated data are divided into train (60%), test (20%) and validation (20 %) and a pre-trained image classification algorithm is fine-tuned for predicting the three image categories. 6) Prediction of all grid cell categories and estimation of precision of algorithm. Output is vector file formatted maps with three distribution categories at a 10 × 10 km resolution. See the "Methods" section for specific details.

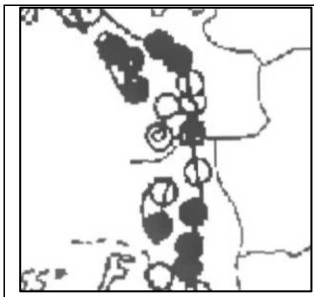
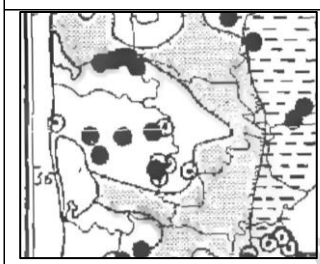
Image classification

To retrieve the species distribution data from the TBU maps (Fig. 1, black-and-white symbols on black-and-white background), we trained an image classification algorithm to recognize the symbols for the different levels of presence and absence in Python version 3.12[70]. We fine-tuned a high performance pre-trained ResNet-18[71] convolutional neural network provided by the PyTorch Python library[72] to predict the classes of the grid cells. The algorithm has been used for a broad range of image classification tasks. In ecology, it has for example been used for computer vision tasks such as coral reef monitoring[73], identification of turtle species from Google images[74] and camera trap identifications[75]. To train the algorithm, we chose 50 out of the 1340 georeferenced maps containing as much variation in symbology as possible. The 50 maps were divided into smaller map parts based on a grid net of 10 × 10 km resolution. We used the extent and projection of the grid net of a recent national atlas survey of Denmark (Atlas Flora Danica[61]). The grid cell resolution of 10 × 10 km was chosen due to the resolution of the maps, in which the real-world area of a dot symbol corresponds to nearly 10 × 10 km. In total, the 50 selected maps amounted to 47865 grid cell images that we annotated into three categories based on presence, abundance and absence (Table 2); 0) absence, background and where species has been found in the past, but has since disappeared, 1) presence; here-and-there, 2) presence; common to very common, actual herbarium and literature records. The annotated grid cell images were randomly divided into training (60 %), validation (20%) and test (20 %) datasets. We trained the model for 25 iterations of the entire training dataset. To evaluate model performance during training, we recorded accuracy and F1 scores for the training and validation dataset. To optimize the model parameters, we used stochastic gradient descent, which updates the model based on small random samples of the data to improve learning efficiency. We also applied a StepLR learning rate scheduler, which reduces the learning rate at fixed intervals to allow more precise parameter updates as training of the model progress. After training of the model, we evaluated the final model on the independent test dataset, using the previously mentioned performance metrics. Finally, we applied the fine-tuned model to classify all grid cells across all maps. All code for the image classification algorithm, training, validation and test datasets are found in the code repository on Zenodo[64].

Overall, the image classification algorithm reached an accuracy score of 98 percent. For each category, the algorithm performed best for predicting the background, which was also overrepresented among categories (absence category 2). Sometimes, multiple symbols overlapped and sometimes maps contained other, more sporadically used, ad hoc symbols, different from the ones mentioned in Table 2, such as triangles and squares representing herbarium specimens or literature mentions. In a few cases, standard symbols were used in non-standard ways depending on the specific map. For example, sometimes an open circle could represent a subspecies' distribution instead of the species being absent. For these cases, and for validating all predictions, we manually cleaned the categories of the maps. Across all maps, we also chose to distinguish grid cells with no records of any species from absence by coding them as NA. Both the annotation and the manual cleaning of the maps were based on the following criteria (Table 3):

Table 3. Image and symbol cases from the TBU maps and the choices made in the annotation process. The same situations and choices were used for manual cleaning of the shapefile maps after prediction of presence and absence categories.

Cases	**Description**	**Choice**
	Black dots or bird's eye dots overlapping more than one grid cell. The example shows four grid cells, where three black dots overlap all, but only count in the grid cells, which overlap the centre of the black dot (blue shade).	Each dot was counted only once, within a single 10 × 10 km grid cell. When a dot overlapped or was predicted to overlap multiple grid cells, we manually assigned it to the grid cell containing the largest proportion of the dot's area. In cases where the overlap was visually ambiguous, we used the QGIS measurement tool to determine the shortest distance from the circle's perimeter to the grid cell borders, assigning the dot to the cell with the greatest area coverage based on these measurements.
	Black dots or bird's eye dots overlapping with dashed horizontal black lines.	If a dot occurred within a grid cell that also contained dashed lines, it was categorized as presence category 1 rather than presence category 2 of "here and there", since dots represent concrete herbarium or field records, while dashed hatching is based on expert judgement.
	Dots, full hatching or dashed hatching in the same grid cell as open circles. In this example, black dots overlap with open dots in the same grid cell shaded blue.	Dots and full hatching were prioritized in classification. Therefore, when a grid cell contained both black dots and open dots (symbol for long vanished occurrences), the cell was categorized as presence.

	Non-standard use of symbols. In this example, open circles, which normally represent vanished occurrences, represent *Spartina × townsendii* agg. in a map otherwise made to show a different taxon, *Spartina alterniflora × maritima*	In some maps, open and black dots were used to denote subspecies rather than their standard usage. In these cases, we did not distinguish between subspecies, but merged them at the species level, reassigning their symbols to the presence category ('1'; Table 2).
	Non-standard symbols. This example shows squares used to represent *Galium pumilum* ssp. *vulgatum*, while the standard dots have been used to represent *Galium pumilum* ssp. *slesvicense*.	A few maps included additional symbols for presence (category '1'; Table 2), such as squares, crosses, or triangles, often used to depict separate taxa, e.g. subspecies, on the same map. These were rarely recognized by the image recognition algorithm due to insufficient training data, and were manually corrected by consulting the original map legends in the TBU publications.
	Uncertain symbols (e.g question marks or shaded areas). In this example, the grey shade represents glaciofluvial plains in Western Jutland, which was used to interpret regional absence of *Galium pumilum* subsp. *septentrionale*.	When symbol interpretation was uncertain or equivocal, we referred to the regional occupancy descriptions provided in the TBU publications to guide classification decisions.

Correction of position of Bornholm

On all TBU maps, the island of Bornholm was placed in an inserted frame in the corner of the general map of Denmark (e.g. Fig. 1, lower right corner)). We corrected this on the categorised shapefiles of the maps by taking the grid cell IDs overlapping Bornholm on the historic maps and moving the values to the grid cell IDs overlapping the geographically true position of Bornholm. The grid cells overlapping the inserted Bornholm did not align completely with the grid cells overlapping the true Bornholm. To align the grid cells, we moved them 3 km east, distorting the grid net for this part of the map but thereby also making direct comparison with modern data from Bornholm possible.

Data Records

This paper presents digitized, spatially explicit maps at a 10 × 10 km resolution of occupancy patterns of 1401 vascular plant taxa (species, subspecies, varieties and a few hybrids) in Denmark, originally published 1931-1980 (Fig. 3). The TBU publications were published from 1931-1976, but we also included the 1980 report on *Rubus* microspecies, which follows the same mapping format as the TBU species maps.

The final cleaned and categorised vector maps and the georeferenced TIFF files of the map images are available from the Figshare repository[69]. We also provide the code used to process the original image maps into georeferenced formats, the image classification algorithm, and the annotations used to predict the symbol categories across all maps. All code, together with a readme file describing each step of the digitization and categorisation workflow, is publicly available on Zenodo[64].

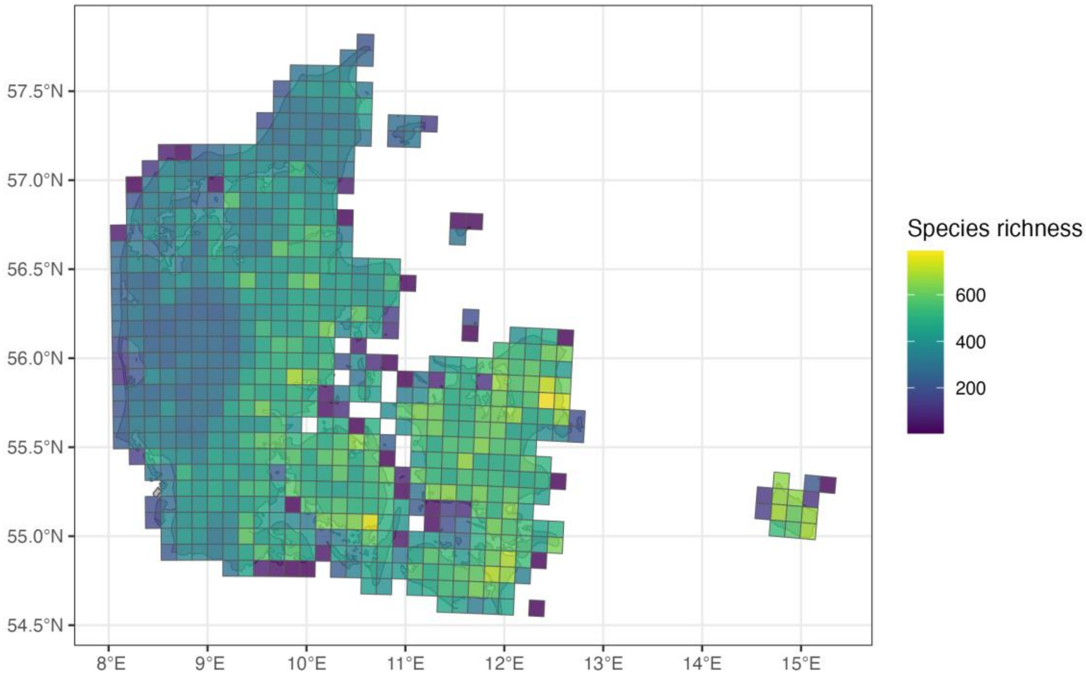


Figure 3. Species richness across Denmark on a 10 × 10 km grid net based on plant species recorded and mapped in the TBU publications. The colour scheme represents species richness distribution from the 1340 TBU maps of 1401 taxa.

Data Overview

Representativity of the Danish flora in the digitized TBU maps

There are approximately 2600 recorded vascular plant species in Denmark[61], including fully established neophytes and casuals found regularly. The TBU publications produced 1340 maps, covering 1401 species, subspecies, varieties, and a few hybrids. Forty-six maps represented two or more subspecies or varieties (using different symbols), while 1294 maps represented one taxon only. In addition, 185 species, including some very common, were treated only in the text of the TBU publications and were not mapped. To assess whether the absence of certain species introduced bias—especially for common, widely distributed species—we compared the species mapped in TBU with the entire Danish flora as defined in the 1981 standard treatise by Hansen et al.[62]. Hansen et al. enumerated 1972 species, subspecies, varieties, and hybrids and their presence and absence across the 57 TBU districts. For comparability, we excluded hybrids and non-naturalized species from this comparison.

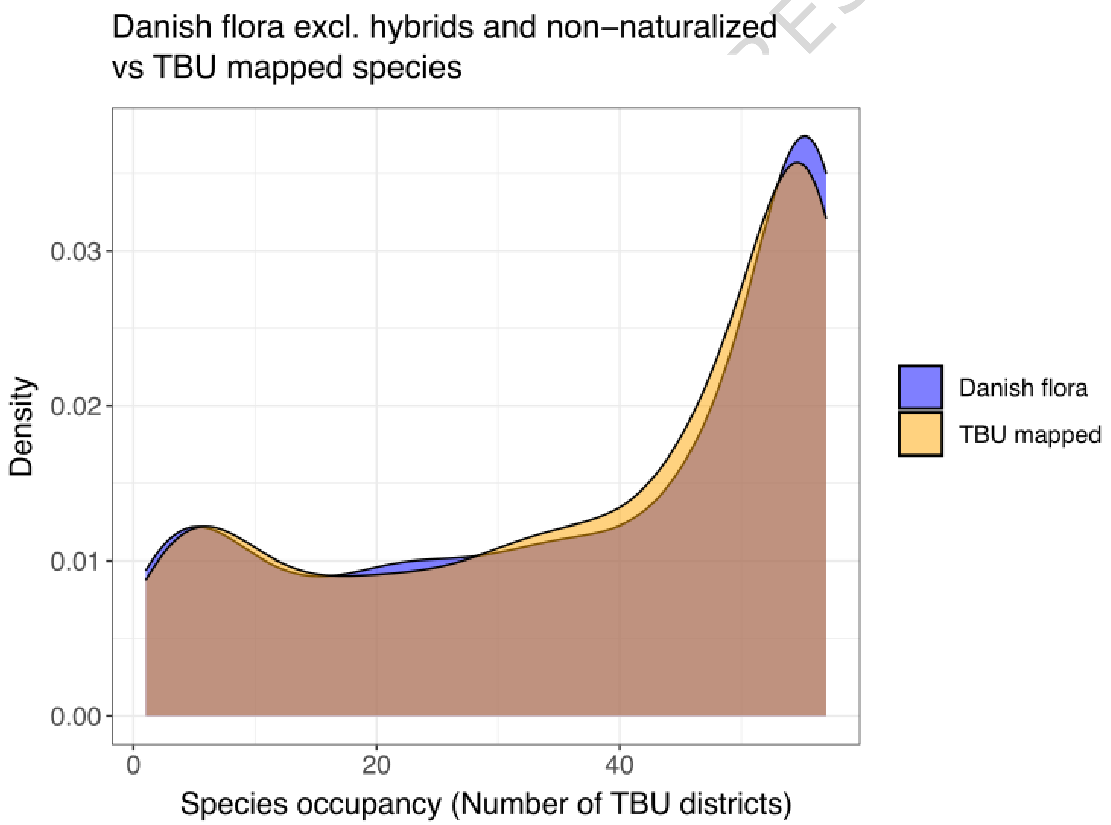


Figure 4. Density plot of species occupancy (calculated as presence across the 57 districts used in the TBU survey) for the mapped species in TBU and for the Danish flora as in Hansen et. al[25]. The purple distribution represents the 1424 species and subspecies included in Hansen et al.[25] (excluding hybrids and non-naturalized species). The orange distribution represents the 1207 species and subspecies mapped in the TBU publications and with occupancy (number of TBU districts) given in Hansen et al[25].

Of the 1401 taxa mapped in the TBU, 1275 species are included in Hansen et al.[62] with 1207 species having a recorded total district occupancy. When comparing these 1207 mapped species with the entire native Danish flora, we observed that very widespread species (i.e., those occurring in more than 50 out of 57 districts) were slightly underrepresented in the TBU maps (Fig. 4). A two-sample Kolmogorov-Smirnov (KS) test confirmed a small, but significant, difference between the two data distributions ($D = 0.05$, $p = 0.012$, $n_1=1207$, $n_2=1424$). However, when including the 185 text-only treated species, the difference disappears (KS test: $D = 0.02$, $p = 0.86$, $n_1=1392$, $n_2=1424$). We interpret the TBU authors' decision to exclude some of the very widespread species from mapping as a recognition that potential gaps in the mapped distributions would be just as likely to reflect limited search effort as true absences of the species, without a reliable way to distinguish between the two. There are 86 species mentioned in Hansen et al.[62] which are neither mapped nor text-treated in TBU; most of these are mentioned in the sections of the TBU publications treating non-naturalized taxa.

The TBU publications include 63 mapped and 73 text-treated taxa that were not considered part of the Danish flora in 1981 by Hansen et al.[62], and lack recorded district occupancy data in that source. A detailed alignment between TBU and Hansen et al.[62] is available on Figshare repository[69].

To compare the distribution of nativeness status of species mapped in TBU with the Danish flora, we used the graduated nativeness system from Colding-Jørgensen et al.[76], which is based on the Euro+Med Plantbase[77]. Species nativeness was categorised as native (native to Denmark according to Euro+Med), near-native (not native, but native to one or more adjacent countries), distantly native (not native, but native to one or more countries further afield in the region covered by Euro+Med) or alien (not native to the Euro+Med area). Species absent from Denmark according to Euro+Med, but present according to the TBU, were manually reclassified from NA to their appropriate category. Species included in Hansen et al.[62] but not mapped by TBU were mainly non-native species (Fig. 5). When text-only treated species were included, the representation of native species became equal between TBU and Hansen et al.[62], and the differences in non-native species categories decreased slightly.

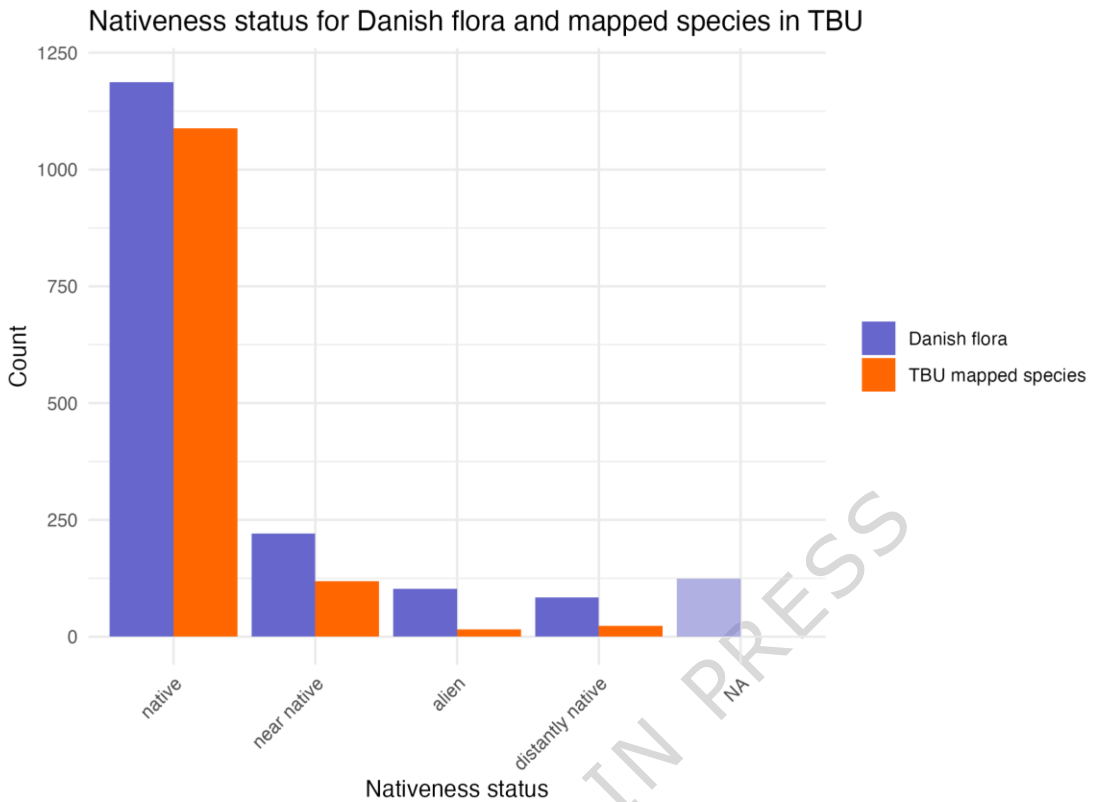


Figure 5. Proportional distribution in nativeness status between species mapped in TBU publications and the entire Danish flora as in Hansen et al.[62]. Most of the species that were not mapped in TBU publications were non-native; most alien species were not mapped in the TBU, the same for distantly native species, while half of the near native species were mapped in the TBU. NA-values in nativeness are species not recognized as part of the Danish flora by the Euro+Med database.

Technical Validation

Digitization validation

We checked 50 random maps and found a maximum of 3 km deviation from the line of the coast on the maps and the real coastline in a few places. Although this is a visible uncertainty, it only occurred in a few places and the deviation was well below the grid size of 10 × 10 km and therefore deemed acceptable.

Taxonomical validation

Species and subspecies taxonomy and nomenclature were standardised and are presented according to World Flora Online Plant List[78], using the Taxonomic Name Resolution Service (TNRS)[79]. We additionally aligned the taxonomy with Danish nomenclature following the recent atlas survey of Atlas Flora Danica[19] and Hansen et al.[25]. The number of recognised taxa varies slightly depending on the taxonomy and nomenclature used. For example, hybrids and most varieties are not recognised in TNRS, but are recognised in the Danish nomenclature system. Consequently, our dataset contains 1239 taxa when matched with the World Flora Online Plant List, but 1315 taxa when aligned with the Danish nomenclature system according to Hansen et al.[25].

Usage Notes

The data set provides a unique resource for analysing historic occupancy patterns of vascular plant species and subspecies in Denmark. Users can combine these data with modern atlas surveys to quantify long-term changes to species' distribution and plant diversity at a 10 × 10 km resolution. Because the taxonomic framework is well defined, the data are suitable for comparative analyses across time and for incorporation into larger synthesis work. In particular, the maps enable users to reconstruct biodiversity patterns prior to the profound land-use changes in the mid-19th century.

To facilitate the use of data across different taxonomic standards, we provide a synonym list that aligns historic taxon names with those used in the most recent Atlas survey for Denmark[61] and with World Flora Online[78]. The list is intended to guide users through changes in species circumscription and nomenclature, and specifies whether species depicted on the TBU maps can be compared directly to taxa as they are defined in modern plant taxonomy. For instance, since the publication of the TBU, a number of species have been acknowledged that were previously perceived as part of the intraspecific variation in more broadly defined species (e.g., *Dryopteris dilatata* and *Dryopteris expansa* were both rendered under the taxon name *Dryopteris dilatata* in the TBU). In a few cases, a single historic map represents the combined occupancy of two species, (e.g. *Sorbus rupicola* and *Sorbus aria*). Users should therefore consult the taxonomy synonym list to determine whether direct comparisons with modern taxa are appropriate or whether aggregation of taxa is required.

The maps include two levels of presence, where one is based on the expert-judged abundance level "here and there". Users should be aware that this presence category reflects expert opinion of coarse-scale density of occurrences and may represent anything from very scattered occurrences, which should result in a checkerboard-like occurrence pattern in the 10 × 10 km grid, to less scattered occurrences resulting in presence throughout a set of contiguous 10 × 10 km grid cells. How to treat this category in analyses is therefore at the discretion of the user. However, we note that, according to Oredsson[80], species with low landscape frequency have detection probabilities at a 10 × 10 km resolution comparable to those of species occurring at high landscape frequency. Consequently, both abundance levels, “common to very common” and “here and there”, can reasonably be interpreted as indicating presence across grid cells at the spatial scale used here.

The maps also include two levels of absence, both of which users should interpret with care. Two distinct situations may lead to a species being coded as absent, and these have different implications. First absence category is based on the focal species being unrecorded, i.e. no presence symbols in the respective grid cell. In some grid cells (e.g. overlapping small islands), a small number of species have been recorded, and the remaining species have not and are therefore scored as absent. It is difficult to determine whether these remaining species were genuinely absent or simply not surveyed, and in such cases the grid cell may be better considered under-sampled rather than all species confirmed absent. We distinguished grid cells with no records of any species across all maps by coding them as NA rather than absent but refrained from doing so in cells containing one record of a species. Grid cells with only a few recorded species may therefore be more appropriately treated as missing data (NA) for all other species in analyses where true absences are required, but we leave this to the user’s discretion. To obtain more reliable absence estimates, users may consider applying methods that correct for uneven sampling effort, ie. Frescalo[81] or similar, which can help distinguish true absences from under-sampled localities.

The other absent category is based on grid cells where a species had been recorded sometime prior to the early-20th-century survey (1904–1923) but was no longer considered present at the time of publication of the maps. The interval between the past record and the disappearance was not standardised across contributors, and no fixed temporal threshold was applied. Instead, contributors noted these cases when a species previously known from their district had ceased to occur by the time of writing. Because the national survey was conducted progressively across districts, this “disappearance” category reflects different local baseline dates rather than one unified national year. For the purpose of reconstructing species’ occupancy at the TBU survey baseline, these “disappeared” localities are treated as absences, consistent with the original authors’ intention. However, users analysing earlier baselines or longer-term presence-history questions may legitimately treat these open-circle cells as historical presences. To enable this flexibility, the original symbol categories are preserved in the dataset so that users may reclassify “disappeared” records as presences, absences, or a separate category according to their analytical needs.

Data Availability

The vector maps and the georeferenced TIFF files of the map images are openly available from the Figshare repository at: https://doi.org/10.6084/m9.figshare.30609671.

Code Availability

A versioned copy of the code used for digitizing and categorising the historic maps with QGIS, Python and R are available in the Zenodo repository at: https://doi.org/10.5281/zenodo.17600480.

Acknowledgements

We are deeply grateful to the numerous people organising and volunteering in the TBU survey. We thank the Danish Botanical Society for giving access to the published maps from their archive. The help by Henrik Ærenlund Pedersen with interpreting Orchidaceae taxonomy from the historic maps to modern taxon circumscription is much appreciated. Leonard Blaschek provided valuable help for writing the macro codes for Fiji. Last, we appreciate the efforts of Nanna Marie Christiansen to clean the predicted categories of the historic maps and help with aligning taxonomy.

Funding

This work received the financial support provided by the NovoNordisk Foundation (SR's starting grant NNF22OC0078703).

Author contributions

HHB, CL and SR conceived the idea and DMC developed it. DMC digitized and georeferenced the maps with help from MA. DMC and TO extracted the occupancy symbols on the maps and assigned them to shapefile grid nets. DMC aligned the taxonomy with help from HHB and CL. DMC led the writing of the manuscript and produced all figures with input from all authors.

Competing interests

The authors declare no conflict of interest.

References

1. Lewis, S. L. & Maslin, M. A. Defining the Anthropocene. *Nature* **519**, 171–180 (2015) https://doi.org/10.1038/nature14258.
2. Tilman, D. *et al.* Future threats to biodiversity and pathways to their prevention. *Nature* **546**, 73–81 (2017) https://doi.org/10.1038/nature22900.
3. Dornelas, M. *et al.* Looking back on biodiversity change: lessons for the road ahead. *Philosophical Transactions of the Royal Society B: Biological Sciences* **378**, 20220199 (2023) https://doi.org/10.1098/rstb.2022.0199.
4. Mihoub, J.-B. *et al.* Setting temporal baselines for biodiversity: the limits of available monitoring data for capturing the full impact of anthropogenic pressures. *Sci Rep* **7**, 41591 (2017) https://doi.org/10.1038/srep41591.
5. Svenning, J.-C., Kerr, M. R., Mungi, N. A., Ordonez, A. & Riede, F. Defining the Anthropocene as a geological epoch captures human impacts' triphasic nature to empower science and action. *One Earth* **7**, 1678–1681 (2024) https://doi.org/10.1016/j.oneear.2024.08.004.
6. Papworth, S. K., Rist, J., Coad, L. & Milner-Gulland, E. J. Evidence for shifting baseline syndrome in conservation. *Conservation Letters* **2**, 93–100 (2009) https://doi.org/10.1111/j.1755-263X.2009.00049.x.
7. Magurran, A. E. *et al.* Long-term datasets in biodiversity research and monitoring: assessing change in ecological communities through time. *Trends in Ecology & Evolution* **25**, 574–582 (2010) https://doi.org/10.1016/j.tree.2010.06.016.

8. MacDougall, A. S. *et al.* Comparison of the distribution and phenology of Arctic Mountain plants between the early 20th and 21st centuries. *Global Change Biology* **27**, 5070–5083 (2021) https://doi.org/10.1111/gcb.15767.
9. Morueta-Holme, N. *et al.* Strong upslope shifts in Chimborazo's vegetation over two centuries since Humboldt. *Proceedings of the National Academy of Sciences* **112**, 12741–12745 (2015) https://doi.org/10.1073/pnas.1509938112.
10. Jandt, U. *et al.* More losses than gains during one century of plant biodiversity change in Germany. *Nature* **611**, 512–518 (2022) https://doi.org/10.1038/s41586-022-05320-w.
11. Blowes, S. A. *et al.* The geography of biodiversity change in marine and terrestrial assemblages. *Science* **366**, 339–345 (2019) https://doi.org/10.1126/science.aaw1620.
12. Blowes, S. A. *et al.* Synthesis reveals approximately balanced biotic differentiation and homogenization. *Science Advances* **10**, eadj9395 (2024) https://doi.org/10.1126/sciadv.adj9395.
13. Staude, I. R. *et al.* Directional turnover towards larger-ranged plants over time and across habitats. *Ecology Letters* **25**, 466–482 (2022) https://doi.org/10.1111/ele.13937.
14. Ostenfeldt, C. H. *The Distribution within Denmark of the Higher Plants. Results of the Topographic-Botanical Investigation. I. A Brief Historical Survey of the Investigation.* Kongelige Danske Videnskabernes Selskabs Skrifter - Naturvidenskabelig og Mathematisk Afdeling, 9. Rk. 3, 1-19. (1931).
15. Pedersen, H. Æ. Tidligere inventering og kortlægning af den danske flora. in *Atlas Flora Danica* vol. 1. Baggrund, metode og analyse 135–150 (Gyldendal, 2015).
16. Jessen, K. *The Distribution within Denmark of the Higher Plants. Results of the Topographic-Botanical Investigation. II. The Distribution of the Papilionaceæ within Denmark.* Kongelige Danske Videnskabernes Selskabs Skrifter - Naturvidenskabelig og Mathematisk Afdeling, 9. Rk. 3, 1-19. (1931).
17. Vestergaard, P., Hansen, K. & Society, D. B. *Opera Botanica : Distribution of vascscular plants 1989. 96 / Peter Vestergaard and Kjeld Hansen*. (Peter Vestergaard et Kjeld Hansen. Copenhagen, 1989).

18. Jessen, K. Liliiflorernes Udbredelse i Danmark. Danmarks Topografisk-Botaniske Undersøgelse nr. 1. *Botanisk Tidsskrift* **43**, 71–132 (1935).

19. Sørensen, T. Primulaceernes Udbredelse i Danmark. Danmarks Topografisk-Botaniske Undersøgelse nr. 2. *Botanisk Tidsskrift* **43**, 133–172 (1935).

20. Böcher, T. W. Udbredelse af Ericaceae, Vacciniaceae og Empetraceae i Danmark. Danmarks Topografisk-Botaniske Undersøgelse nr. 3. *Botanisk Tidsskrift* **44**, 5–35 (1937) .

21. Wiinstedt, K. Juncaceernes Udbredelse i Danmark. Danmarks Topografisk-Botaniske Undersøgelse nr. 4. *Botanisk Tidsskrift* **44**, 41–125 (1937).

22. Grøntved, J. Polygonaceernes Udbredelse i Danmark. Danmarks Topografisk-Botaniske Undersøgelse nr. 5. *Botanisk Tidsskrift* **45**, 10–58 (1939).

23. Køie, A. Hypericaceernes Udbredelse i Danmark. Danmarks Topografisk-Botaniske Undersøgelse nr. 6. *Botanisk Tidsskrift* **45**, 59–72 (1939).

24. Køie, A. & Køie, M. Udbredelsen af Geraniaceæ, Araceæ, Lemnaceæ og Droseraceæ i Danmark. Danmarks Topografisk-Botaniske Undersøgelse nr. 7. *Botanisk Tidsskrift* **45**, 73–107 (1939).

25. Wiinstedt, K. Hieraciumarternes Udbredelse i Danmark. Danmarks Topografisk-Botaniske Undersøgelse nr. 8. *Botanisk Tidsskrift* **45**, 101–129 (1939).

26. Wiinstedt, K. Cyperaceernes Udbredelse i Danmark. I. Scirpoideae. Danmarks Topografisk-Botaniske Undersøgelse nr. 9. *Botanisk Tidsskrift* **47**, 3–64 (1943).

27. Mikkelsen, V. M. Udbredelse af Juncaginaceae, Alismataceae og Hydrocharitaceae i Danmark. Danmarks Topografisk-Botaniske Undersøgelse nr. 10. *Botanisk Tidsskrift* **47**, 65–93 (1943).

28. Hoff, M. Crassulaceernes og Saxifragaceernes Udbredelse i Danmark. Danmarks Topografisk-Botaniske Undersøgelse nr. 11. *Botanisk Tidsskrift* **47**, 95–121 (1943).

29. Andersen, A. Pyrolaceernes og Plumbaginaceernes Udbredelse i Danmark. Danmarks Topografisk-Botaniske Undersøgelse nr. 12. *Botanisk Tidsskrift* **47**, 123-142. (1943).

30. Wiinstedt, K. Cyperaceernes Udbredelse i Danmark. II. Caricoideae. Danmarks Topografisk-Botaniske Undersøgelse nr. 13. *Botanisk Tidsskrift* **47**, 143–244 (1945).

31. Hansen, A. Campanulaceernes og Lobeliaceernes Udbredelse i Danmark. Danmarks Topografisk-Botaniske Undersøgelse nr. 14. *Botanisk Tidsskrift* **47**, 245–276 (1948).

32. Grøntved, J. Orchideernes Udbredelse i Danmark. Danmarks Topografisk-Botaniske Undersøgelse nr. 15. *Botanisk Tidsskrift* **47**, 277–370 (1948).

33. Egholm, B. Umbelliferernes udbredelse i Danmark. Danmarks Topografisk-Botaniske Undersøgelse nr. 16. *Botanisk Tidsskrift* **47**, 373–480 (1951).

34. Hansen, A. Udbredelsen af Caprifoliaceae, Adoxaceae, Dipsacaceae og Cucurbitaceae i Danmark. Danmarks Topografisk-Botaniske Undersøgelse nr. 17. *Botanisk Tidsskrift* **47**, 481–509 (1951).

35. Wiinstedt, K. Pteridophyternes Udbredelse i Danmark. Danmarks Topografisk-Botaniske Undersøgelse nr. 18. *Botanisk Tidsskrift* **49**, 305–388 (1953).

36. Rasmussen, S. M. Euphorbiaceernes, Malvaceernes og Violaceernes udbredelse i Danmark. Danmarks Topografisk-Botaniske Undersøgelse nr. 20. *Botanisk Tidsskrift* **50**, 239–278 (1954).

37. Grøntved, J. Typhaceernes og Sparganiaceernes Udbredelse i Danmark. Danmarks Topografisk-Botaniske Undersøgelse nr. 19. *Botanisk Tidsskrift* **50**, 209–238 (1954).

38. Pedersen, A. Rubiaceernes, Polygalaceernes, Linaceernes, Oxalidaceernes og Balsaminaceernes udbredelse i Danmark. Danmarks Topografisk-Botaniske Undersøgelse nr. 21. *Botanisk Tidsskrift* **53**, 139–196 (1956).

39. Larsen, K. Ranunculaceernes udbredelse i Danmark. Danmarks Topografisk-Botaniske Undersøgelse nr. 22. *Botanisk Tidsskrift* **53**, 197–252 (1956).

40. Pedersen, A. Cruciferernes udbredelse i Danmark. Danmarks Topografisk-Botaniske Undersøgelse nr. 23. *Botanisk Tidsskrift* **54**, 191–304 (1958).

41. Hansen, A. Gentianaceernes, Menyathaceernes, Asclepiadaceernes og Apocynaceernes udbredelse i Danmark. Danmarks Topografisk-Botaniske Undersøgelse nr. 24. *Botanisk Tidsskrift* **54**, 305–332 (1958).

42. Pedersen, A. Caryophyllaceernes udbredelse i Danmark. Danmarks Topografisk-Botaniske Undersøgelse nr. 25. *Botanisk Tidsskrift* **55**, 157–267 (1959).

43. Hansen, A. Plantaginaceernes og Lentibulariaceernes udbredelse i Danmark. Danmarks Topografisk-Botaniske Undersøgelse nr. 26. *Botanisk Tidsskrift* **56**, 1–35 (1960).

44. Larsen, K. & Pedersen, A. Papaveraceernes, Fumariaceernes, Nymphaeaceernes, Ceratophyllaceernes, Elatinaceernes, Halorrhagidaceernes, Hippuridaceernes og Lythraceernes udbredelse i Danmark. Danmarks Topografisk-Botaniske Undersøgelse nr. 27. *Botanisk Tidsskrift* **56**, 37–86 (1960).

45. Pedersen, A. Kurvblomsternes udbredelse i Danmark. Danmarks Topografisk-Botaniske Undersøgelse nr. 28. *Botanisk Tidsskrift* **57**, 81–289 (1961).

46. Pedersen, A. Scrophulariaceernes og Orobanchaceernes udbredelse i Danmark. Danmarks Topografisk-Botaniske Undersøgelse nr. 29. *Botanisk Tidsskrift* **59**, 1–140 (1963).

47. Hansen, A. Convolvulaceernes, Cuscutaceernes, Hydrophyllaceernes, Polemoniaceernes og Solanaceernes udbredelse i Danmark. Danmarks Topografisk-Botaniske Undersøgelse nr. 30. *Botanisk Tidsskrift* **59**, 141–176 (1963).

48. Rasmussen, S. M. Boraginaceernes udbredelse i Danmark. Danmarks Topografisk-Botaniske Undersøgelse nr. 31. *Botanisk Tidsskrift* **60**, 261–315 (1965).

49. Pedersen, A. Rosaceernes udbredelse i Danmark, I. Underfamilierne Spiraeoideae, Dryadoideae og Rosoideae eksklusive Rubus fruticosus coll., Rubus caesius coll. og Rubus corylifolius coll. Danmarks Topografisk-Botaniske Undersøgelse nr. 32. *Botanisk Tidsskrift* **61**, 145–270 (1965).

50. Pedersen, A. Cannabaceernes, Urticaceernes, Santalaceernes, Aristolochiaceernes, Resedaceernes og Cistaceernes udbredelse i Danmark. Danmarks Topografisk-Botaniske Undersøgelse nr. 33. *Botanisk Tidsskrift* **62**, 85–122 (1966).

51. Pedersen, A. Callitrichaceernes udbredelse i Danmark. Danmarks Topografisk-Botaniske Undersøgelse nr. 34. *Botanisk Tidsskrift* **62**, 123–145 (1966).

52. Hansen, A. & Pedersen, A. Chenopodiaceernes og Amaranthaceernes udbredelse i Danmark. Danmarks Topografisk-Botaniske Undersøgelse nr. 35. *Botanisk Tidsskrift* **63**, 205–288 (1968).

53. Ødum, S. Udbredelsen af træer og buske i Danmark. Danmarks Topografisk-Botaniske Undersøgelse nr. 36. *Botanisk Tidsskrift* **64**, 1–118 (1968).

54. Pedersen, A. Labiaternes og Verbenaceernes udbredelse i Danmark. Danmarks Topografisk-Botaniske Undersøgelse nr. 37. *Botanisk Tidsskrift* **64**, 285–379 (1969).

55. Jensen, N. Oenotheraceernes udbredelse i Danmark. Danmarks Topografisk-Botaniske Undersøgelse nr. 38. *Botanisk Tidsskrift* **66**, 137–170 (1971).

56. Pedersen, A. Gramineernes udbredelse i Danmark. Spontane og naturaliserede arter. Danmarks Topografisk-Botaniske Undersøgelse nr. 39a. *Botanisk Tidsskrift* **68**, 177–343 (1974).

57. Pedersen, A. Najadaceernes, Potomogetonaceernes, Ruppiaceernes, Zannichelliaceernes og Zosteraceernes udbredelse i Danmark. Danmarks Topografisk-Botaniske Undersøgelse nr. 40. *Botanisk Tidsskrift* **70**, 203–262 (1976).

58. Hansen, A. & Pedersen, A. Portulacaceernes og Valerianaceernes udbredelse i Danmark. Danmarks Topografisk-Botaniske Undersøgelse nr. 41. *Botanisk Tidsskrift* **71**, 57–74 (1976).

59. Pedersen, A. Rosaceernes udbredelse i Danmark, II. Subgenus Rubus Sect. Rubus, Sect. Corylifolii og Sect. Caesii. Danmarks Topografisk-Botaniske Undersøgelse nr. 42. *Botanisk Tidsskrift* **75**, 1–50 (1980).

60. Wind, P. & Pihl, S. Den danske rødliste 2019. Aarhus Universitet, DCE - Nationalt Center for Miljø og Energi (2019).

61. Hartvig, P. & Vestergaard, P. *Atlas Flora Danica*. (Gyldendal, København, 2015).

62. Hansen, K. *Dansk feltflora*. (Gyldendal Boghandel Nordisk Forlag, Copenhagen, 1981).

63. Schindelin, J. *et al.* Fiji: an open-source platform for biological-image analysis. *Nat Methods* **9**, 676–682 (2012) https://doi.org/10.1038/nmeth.2019.

64. Christiansen, D. M. Code for From paper to pixels - Digitized maps of vascular plant species distributions in Denmark in the early 20th century. https://doi.org/10.5281/zenodo.17600480.

65. QGIS Development Team. QGIS Geographic Information System. (2023).

66. R Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. <https://www.R-project.org/>. (2024).

67. Warnes, G. R., Bolker, B., Lumley, T. & Warnes, M. G. R. Package ‘gtools’. R Package version Version 3.9.5, 1–52. https://github.com/r-gregmisc/gtools (2023).

68. O’Brien, J. gdalUtilities: Wrappers for “GDAL” Utilities Executables. R package version 1.2.5 https://CRAN.R-project.org/package=gdalUtilities (2023).

69. Christiansen, D. M. *et al.* Digitized maps of vascular plant species distributions in Denmark in the early 20th century. Figshare. https://doi.org/10.6084/m9.figshare.30609671.

70. Python Software Foundation. Python Language Reference, version 3.12. Available at http://www.python.org.

71. He, K., Zhang, X., Ren, S. & Sun, J. Deep Residual Learning for Image Recognition. in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* 770–778 (2016). doi:10.1109/CVPR.2016.90.

72. Paszke, A. et al. PyTorch: an imperative style, high-performance deep learning library. In Advances in Neural Information Processing Systems 32 (eds Wallach, H. M. et al.) 8026–8037 (ACM, 2019).

73. Marre, G. *et al.* Deep convolutional neural networks to monitor coralligenous reefs: Operationalizing biodiversity and ecological assessment. *Ecological Informatics* **59**, 101110 (2020) https://doi.org/10.1016/j.ecoinf.2020.101110.

74. Baek, J.-W., Kim, J.-I. & Kim, C.-B. Deep learning-based image classification of turtles imported into Korea. *Sci Rep* **13**, 21677 (2023) https://doi.org/10.1038/s41598-023-49022-3.

75. Willi, M. *et al.* Identifying animal species in camera trap images using deep learning and citizen science. *Methods in Ecology and Evolution* **10**, 80–91 (2019) https://doi.org/10.1111/2041-210X.13099.

76. Colding-Jørgensen, C. T., Ejrnæs, R., Svenning, J.-C. & Bruun, H. H. A graduated nativeness definition for overcoming dilemmas and difficulties of vascular plant species. *Oikos* **2026**, e11635 (2026) https://doi.org/10.1002/oik.11635.

77. Euro+Med 2006+ [continuously updated]: Euro+Med PlantBase – the information resource for Euro-Mediterranean plant diversity. (2025).

78. The WFO Plant List | World Flora Online. (2025) https://wfoplantlist.org/.

79. Boyle, B. *et al.* The taxonomic name resolution service: an online tool for automated standardization of plant names. *BMC Bioinformatics* **14**, 16 (2013) https://doi.org/10.1186/1471-2105-14-16.

80. Oredsson, A. Frekvensskattning som en funktion av närvaro och frånvaro. *Svensk Botanisk Tidskrift* **75**, 109–113 (1981) .

81. Hill, M. O. Local frequency as a key to interpreting species occurrence data when recording effort is not known. *Methods in Ecology and Evolution* **3**, 195–205 (2012) https://doi.org/10.1111/j.2041-210X.2011.00146.x.