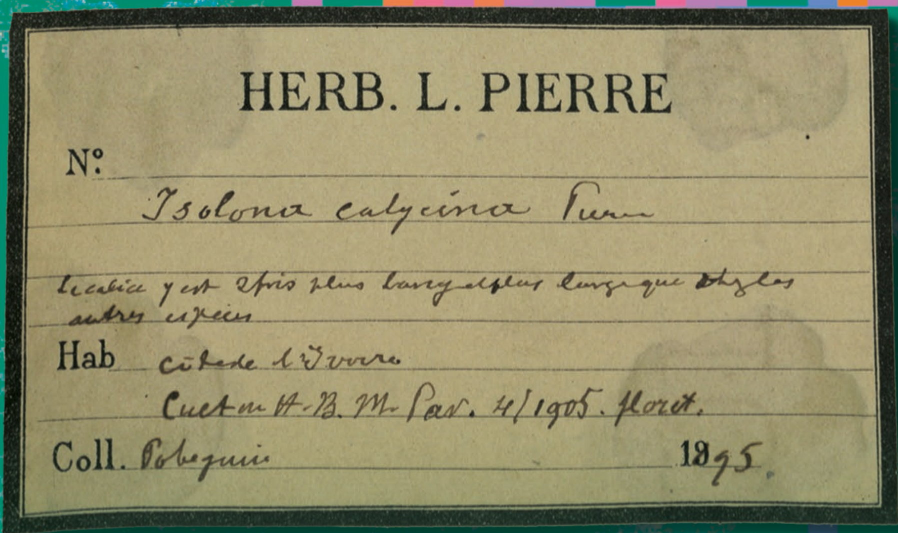
adansonia

2024 • 46 • 18

High-quality herbarium-label transcription by citizen scientists improves taxonomic and spatial representation of the tropical plant family Annonaceae

Serafin J. R. STREIFF, Erica O. RAVOMANANA,
Mijoro RAKOTOARINIVO, Marc PIGNAL,
Eva PEREZ PIMPARÉ, Roy H. J. ERKENS &
Thomas L. P. COUVREUR



art. 46 (18) — Published on 4 November 2024
www.adansonia.com

PUBLICATIONS SCIENTIFIQUES
MUSÉUM NATIONAL D'HISTOIRE NATURELLE

Couverture / *Cover*:
Réalisée à partir des Figures de l'article/*Made from the Figures of the article*.

Adansonia est indexé dans / Adansonia *is indexed in:*
– Science Citation Index Expanded (SciSearch®)
– ISI Alerting Services®
– Current Contents® / Agriculture, Biology, and Environmental Sciences®
– Scopus®

Adansonia est distribué en version électronique par / Adansonia *is distributed electronically by:*
– BioOne® (http://www.bioone.org)

Adansonia est une revue en flux continu publiée par les Publications scientifiques du Muséum, Paris
Adansonia *is a fast track journal published by the Museum Science Press, Paris*

Les Publications scientifiques du Muséum publient aussi / *The Museum Science Press also publish:* ***Geodiversitas***, ***Zoosystema***, ***Anthropozoologica***, ***European Journal of Taxonomy***, ***Naturae***, ***Cryptogamie*** **sous-sections** ***Algologie***, ***Bryologie***, ***Mycologie***, ***Comptes Rendus Palevol***

Diffusion – Publications scientifiques Muséum national d'Histoire naturelle
CP 41 – 57 rue Cuvier F-75231 Paris cedex 05 (France)
Tél.: 33 (0)1 40 79 48 05 / Fax: 33 (0)1 40 79 38 40
diff.pub@mnhn.fr / http://sciencepress.mnhn.fr

ISSN (imprimé / *print*): 1280-8571/ ISSN (électronique / *electronic*): 1639-4798

High-quality herbarium-label transcription by citizen scientists improves taxonomic and spatial representation of the tropical plant family Annonaceae

Serafin J. R. STREIFF
DIADE, Université de Montpellier, CIRAD, IRD, Montpellier (France)
sjrstreiff@gmail.com

Erica O. RAVOMANANA
Mijoro RAKOTOARINIVO
Mention Biologie et Écologie végétales, Faculté des Sciences,
Université d'Antananarivo, Antananarivo BP 906101 (Madagascar)

Marc PIGNAL
Institut de Systématique, Evolution, Biodiversité (ISYEB), Muséum national d'Histoire naturelle,
CNRS, Sorbonne Université, EPHE, Université des Antilles,
case postale 39, 57 rue Cuvier, 75231 Paris cedex 05 (France)

Eva PEREZ PIMPARÉ
Infrastructure de Recherche Récolnat, Direction générale déléguée aux Collections,
Muséum national d'Histoire naturelle, 57 rue Cuvier, 75231 Paris cedex 05 (France)

Roy H. J. ERKENS
Maastricht Science Programme, Faculty of Science and Engineering,
Maastricht University, Maastricht (The Netherlands)
and System Earth Science, Faculty of Science and Engineering,
Maastricht University, Venlo (The Netherlands)

Thomas L. P. COUVREUR
DIADE, Université de Montpellier, CIRAD, IRD, Montpellier (France)

Submitted on 23 January 2024 | accepted on 7 May 2024 | published on 4 November 2024

Streiff S. J. R., Ravomanana E. O., Rakotoarinivo M., Pignal M., Perez Pimparé E., Erkens R. H. J. & Couvreur T. L. P. 2024. — High-quality herbarium-label transcription by citizen scientists improves taxonomic and spatial representation of the tropical plant family Annonaceae. *Adansonia*, sér. 3, 46 (18): 173-185. https://doi.org/10.5252/adansonia2024v46a18. http://adansonia.com/46/18

ABSTRACT
Herbarium specimens provide an important and central resource for biodiversity research. Making these records digitally available to end-users represents numerous challenges, in particular, transcribing metadata associated with specimen labels. In this study, we used the citizen science initiative 'Les Herbonautes' and the Récolnat network to transcribe specific data from all herbarium specimen labels stored at the Muséum national d'Histoire naturelle in Paris of the large tropical plant family Annonaceae. We compared this database with publicly available global biodiversity repository data and expert checklists. We investigated spatial and taxonomic advances in data availability at the global and country scales. A total of 20 738 specimens were transcribed over the course of more than two years contributing to and significantly extending the previously available specimen and species data for Annonaceae worldwide. We show that several regions, mainly in Africa and South East Asia not

KEY WORDS
Annonaceae,
participatory transcription,
GBIF,
conservation,
ecology,
Madagascar.

covered by online global datasets, are uniquely available in the P herbarium, probably linked to past history of the museum's botanical exploration. While acknowledging the challenges faced during the transcription of historic specimens by citizen scientists, this study highlights the positive impact of adding records to global datasets both in space and time. This is illustrative for researchers, collection managers, policy makers as well as funders. These datasets will be valuable for numerous future studies in biodiversity research, including ecology, evolution, conservation and climate change science.

RÉSUMÉ

La transcription de haute qualité d'étiquettes d'herbiers par des scientifiques citoyens améliore la représentation taxonomique et spatiale d'une famille de plantes tropicales, les Annonaceae.

Les spécimens d'herbier constituent une ressource importante et centrale pour la recherche sur la biodiversité. Rendre ces données accessibles sous forme numérique aux utilisateurs représente de nombreux défis, en particulier la transcription des métadonnées associées aux étiquettes des échantillons. Dans cette étude, nous avons utilisé le projet de science participative « Les Herbonautes » et le réseau Récolnat pour retranscrire les données spécifiques de toutes les étiquettes de spécimens d'herbier contenues au Muséum national d'Histoire naturelle à Paris d'une famille de plantes tropicales, les Annonaceae. Nous avons comparé cette base de données avec les données de biodiversité en libre accès et des listes taxonomiques expertes. Nous avons étudié l'apport de ces spécimens nouvellement transcrits au niveau spatial et taxonomique à l'échelle mondiale et nationale. Au total, 20 738 spécimens ont été transcrits sur une période de plus de deux ans, contribuant ainsi à enrichir et à étendre considérablement les données précédemment disponibles sur les spécimens et les espèces d'Annonacées au niveau mondial. Nous montrons que plusieurs régions, principalement en Afrique et en Asie du Sud-Est, non couvertes par les données en ligne, sont uniquement disponibles dans l'herbier P, probablement liées à l'histoire de l'exploration botanique du Muséum. Tout en reconnaissant les défis rencontrés lors de la transcription de spécimens historiques par les scientifiques citoyens, cette étude souligne l'impact positif de l'ajout de spécimens aux données globales à la fois aux niveaux spatial et temporel. Ceci est révélateur pour les chercheurs, les gestionnaires de collections, les décideurs politiques ainsi que les bailleurs de fonds. Ces ensembles de données seront précieux pour de nombreuses études futures dans le cadre de la recherche sur la biodiversité, notamment sur l'écologie, l'évolution, la conservation et sur l'impact du changement climatique sur la biodiversité.

MOTS CLÉS
Annonaceae, transcription participative, GBIF, conservation, écologie, Madagascar.

INTRODUCTION

Natural history collections are the foundation of much of our basic knowledge on biodiversity (Krishtalka *et al.* 2016). For plants specifically, herbaria contain vast amounts of information via the storage and curation of herbarium specimens. These collections are central for biodiversity research, specifically taxonomy, and as reference vouchers, both for morphological and molecular data (Turland *et al.* 2018; Davis 2023). Beyond taxonomic considerations, herbarium specimens and their associated information available on the labels provide data on species' ecology (e.g. with isotopic compositions, McLauchlan *et al.* 2010), their phenology (e.g. Willis *et al.* 2017; Gallinat *et al.* 2018; Park *et al.* 2023a), ethnopharmacology (Eloff 1999) and ethnobotanical history (Souza & Hawkins 2017; Van Andel *et al.* 2022), as well as location and timing of their presence (Albani Rocchetti *et al.* 2021; Heberling *et al.* 2019; Davis 2023).

Herbarium specimens were traditionally only accessible on site at respective institutions (Figueiredo & Smith 2010) allowing convenient access to a large number of generally curated specimens. The strong impact of colonial history on botanical collecting has led to many collections originating from tropical regions being concentrated predominantly in western European and North-American herbaria (Park *et al.* 2023a), often home to millions of specimens. The result is that working with these specimens generally goes hand in hand with large logistical effort, including either travel and lodging costs or lengthy loan requests (which also involve certain risks, e.g. https://www.science.org/content/article/botanists-fear-research-slowdown-after-priceless-specimens-destroyed-australian-border).

Development of mass digitisation, i.e., specimen scanning, and databasing, i.e., label information transcription, are being undertaken to make specimens broadly available and accessible (e.g., Thiers *et al.* 2016; Hedrick *et al.* 2020; Hidalga *et al.* 2020; Santos *et al.* 2020). In digital as well as in physical form, the real value of any herbarium specimen not only consists of the dried plant material. The associated collection metadata in form of plant descriptions, observations and locality information on the label contain data equally as valuable. Imaging a physical specimen is relatively straightforward with the appropriate equipment, but transcribing a label and georeferencing the collection locality of the record can be complex and time consuming. Printed or typewritten labels only became common from the 1930s onwards, and are

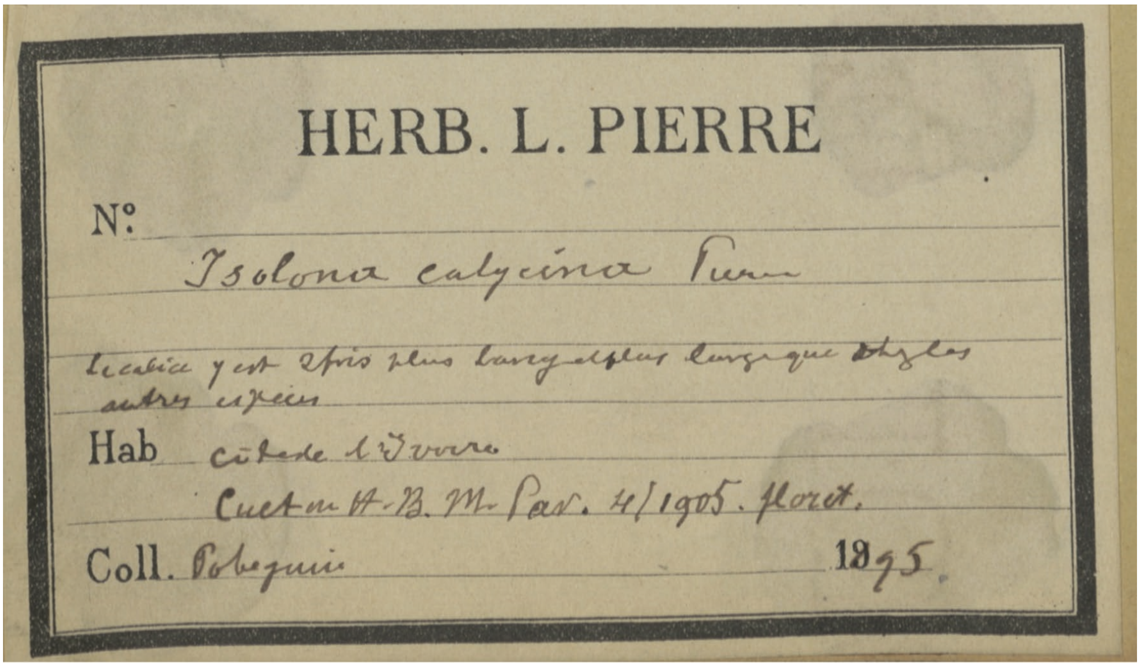


Fig. 1. — Label of specimen P02032119, illustrating some challenges of deciphering specimen labels. Collected by Pobéguin in 1895. No s.n., *Isolona calycina* Pierre, *nomen nudum* in herb. (never validly published). "Le calice y est 2 fois plus longue et plus large que chez les autres espèces" [*The calyx here is two times longer and wider than in the other species*]. "Hab. Côte de l'Ivoire. Cult in H[ortus] B[otanicus] M[usei] Par[isiensis] 4/1905. Floret [blossoms]" (referring to the sampling of the specimen in 1905).

presently still not ubiquitous. Furthermore, labels are not always standardised, and are sometimes nearly indecipherable even to trained readers (e.g., Fig. 1). Consequently, important information is either erroneously – or not at all - interpreted. This includes basic information such as collector names, collection numbers or collection locality. Location names might be either missing, imprecise, or may have changed in the decades or centuries since the date of collecting, complicating georeferencing. Deciphering and transcribing of such labels is done on a daily basis by curators and researchers in herbaria. With the up-and-coming momentum of digitisation requiring accelerating transcription efforts to keep up with imaging, this job has become increasingly intractable (Corlett 2023). Thus, two general strategies of mass digitisation have been taken: a) all specimens are imaged, transcribed and databased in one concerted effort but for a reduced number of specimens; or b) specimens are imaged upfront with minimal metadata for a large number of specimens with transcription and full databasing taking place at a later point in time. In some cases, label data have been transcribed by researchers during taxonomic revisions prior and independent to digitization.

Large scale transcription of label information can be separated in two main strategies: crowdsourcing, either commercial or via 'citizen science' projects (Ridge 2013; Drinkwater *et al.* 2014; Groom *et al.* 2019) or automated, using for example Optical Character Recognition (OCR) (Carranza-Rojas *et al.* 2017; Engledow *et al.* 2018; Kirchhoff *et al.* 2018; Sweeney *et al.* 2018; Dillen *et al.* 2019). Automated approaches have the advantage of being able to process large amounts of data quickly, however, their validation is critical to avoid systematic mistakes (Corlett 2023), even with the more recent development of powerful Large Language Models (Weaver *et al.* 2023). This aspect has led to intermediate approaches where data validation has been part of crowdsourcing projects (Drinkwater *et al.* 2014).

Outsourcing transcription relies on manual entry of label information either via paid services or using citizen science. This process is more time intensive, but has the added advantage that each label transcription can be entered several times, to cross validate the resulting data. Beyond the practical aspect of generating data, and in the spirit of true citizen science, in a natural history collection, this also offers the chance of educating the participants in aspects of biodiversity science and natural history collections (Silvertown 2009).

Irrespective of the way data are generated, large-scale digitisation efforts produce significant amounts of data, which need to be findable, accessible, interoperable and reusable for highest impact and use (i.e. FAIR, Wilkinson *et al.* 2016). Creating easily accessible databases of digitised specimens in such a way can be extremely valuable for both the supplying institution and end-users, who can view, extract information and identify or verify specimens, feeding back to the institution and helping it to keep databases up to date. Institutions can enlarge their outreach and visibility by providing digital specimens and information.

In parallel, the need to facilitate access to biodiversity data, allowing collections to be kept up to date and used is highlighted by estimates that for tropical plant groups, more than half of the specimens might be wrongly identified or filed under an outdated taxonomic name (Goodwin *et al.* 2015). This has important implications for study results, conservation actions and policy decisions (Vogel Ely *et al.* 2017). As a consequence, a number of tropical plant groups remain underexplored, thus species numbers and circumscriptions are expected to change once they are the focus of taxonomic revisions (Cheek *et al.* 2020). An illustrative example is the major tropical family Annonaceae, with about 2 500 species (Nge *et al.* in press). Efforts to describe the diversity of this family have been ongoing for several decades (Maas 1984; Couvreur *et al.* 2012; Erkens *et al.* 2017; Meade & Parnell 2018; Hoekstra *et al.* 2021). Due to their important diversity and ecological role in tropical rain forests (Erkens *et al.* 2023), Annonaceae is used as a proxy to study this biome's evolution (Couvreur *et al.* 2011). Over the years comprehensive taxonomic accounts of most genera (e.g., see the overview of Neotropical revisions in Erkens *et al.* 2017) and multiple regional floras (e.g., Couvreur *et al.* 2022; Maas *et al.* 2023) have been published, together with family-level molecular phylogenies (Chatrou *et al.* 2012; Guo *et al.* 2017; Couvreur *et al.* 2019; Nge *et al.* in press), making Annonaceae one of the taxonomically better-known tropical families. However, to date, for most Annonaceae the spatial data to understand the distribution and ecology of all species is still lacking (Erkens *et al.* 2023).

The Muséum national d'Histoire naturelle in Paris, France (acronym P, Thiers, continuously updated) contains the largest herbarium collection globally, with more than 8 million specimens (Le Bras *et al.* 2017) accumulated over more than 350 years. As for most families, P is a major hub for studying Annonaceae herbarium specimens. However, while more than 99% of the vascular plant specimens of P are now imaged, only a part of the associated label information has been transcribed (estimated at approx. 20%). Transcription and geo-referencing is performed within the project 'Les Herbonautes', based on participatory science through a dedicated web-based portal (http://lesherbonautes.mnhn.fr/) (Rouhan *et al.* 2016; Le Bras *et al.* 2017).

Within an ongoing large-scale project focusing on Annonaceae, we focused on using the 'Les Herbonautes' platform to transcribe and georeference all images of Annonaceae specimens deposited at the P herbarium. This transcription effort provides an excellent opportunity to quantify the value of these large-scale digitisation projects, and how the produced digitised records contribute to our understanding of biodiversity patterns, in particular the tropics, when compared to readily available open access databases such as GBIF. We investigated how transcribing label metadata of *c.* 20 000 Annonaceae herbarium specimens within the 'Les Herbonautes' framework impacted our understanding of the spatio-temporal as well as taxonomic distribution of records on a global scale for this family. We highlight challenges faced during the transcription of specimens by citizen scientists from these images and discuss some limitations of these large datasets and how these might be addressed.

MATERIAL AND METHODS

The collections of the Paris herbarium were imaged in a mass digitisation project from 2010 to 2012. The digitisation process is described in detail in Le Bras *et al.* (2017). Briefly, specimens were imaged by a private contractor on 3 conveyor belt systems in parallel, each with 5 technicians. With this setup, around 6 Million specimens were imaged and minimally databased (accession number, geographic region and taxon name) at an average rate of around 10 000 images a day. For this 'Les Herbonautes' project, a total of 20 738 images of specimens filed under Annonaceae were available in Paris (P). We use the term specimen to refer to (photographed) barcoded sheets, distinct from a gathering (i.e. a single collection event). The distinction between gatherings as signified by collector name and collection number is complicated by some prolific collectors (e.g., Louis Pierre) using a single number for multiple gatherings of (supposedly) same species in different localities.

A small number of the available specimens came from associated herbaria (CHE [1], GUAD [34], MPU [1]) who belong to the Récolnat network (https://www.recolnat.org/fr/). The databasing was separated into five citizen science 'episodes' ("*Bienvenue dans la jungle" – Épisode 1-5*/"Welcome to the jungle" – Episode 1-5), each episode corresponding to one geographical region following how the P herbarium specimens are organized (Le Bras *et al.* (2017): fig. 5). Episodes 1 through 5 focused on Madagascar, Asia, Oceania, the Americas and Africa, respectively.

Citizen scientists were asked to enter data as follows: date collected, collector name, co-collectors, collector number, determined by, country, region, locality (as on label) and georeferenced coordinates. Family, genus and specific epithet had already been entered during the original digitisation process. In addition, we asked two extra questions related to phenology (if the specimen was in flower and/or in fruit) and what the colours were of the flowers and/or fruits as described in the label. Results of these later questions are not reported here. Thus, not all label information was entered (for example identification date; habitat of specimen; ecology). This was a deliberate choice as more data entered lead to a significant increase in time to validate every entry.

For every specimen to be validly databased, all metadata fields as indicated above needed to be recorded at least twice and identically. Discrepancies between transcriptions were flagged, and sent back to be revised by the citizen scientists, and if no resolution was found, the discrepancies were corrected, verified and validated by experts. Each episode remained open until 100% of specimens were validated and conflicts resolved. The resulting dataset is referred to here as the Herbonautes dataset.

To compare the Herbonautes dataset to previously available data, we downloaded all available Annonaceae records from

GBIF (extracting only 'preserved specimens'; accessed 02 March 2023; https://doi.org/10.15468/dl.zamdu6; N.B.: the new Paris dataset had not been submitted at the time of accessing GBIF).

We also compared the Herbonautes dataset to a near complete inventory of Annonaceae in Madagascar using an unpublished dataset of expert-curated specimens of Malagasy Annonaceae (Ravomanana *et al.* pers. comm.). This dataset was generated by compiling specimen information from different resources such as GBIF, Episode 1 of the "Les Herbonautes", Naturalis Biodiversity Center (bioportal), Missouri Botanical Garden (Tropicos®), and the Antananarivo herbarium (Parc Botanique et Zoologique de Tsimbazaza; TAN). It represents a near complete dataset of all Annonaceae specimens from Madagascar, and the best estimate of its diversity to date.

To standardise comparisons between datasets, specimens with missing species level identification and/or spatial information (coordinates) were removed. A custom python script was used to update taxonomy using Plants of the World Online (POWO; https://powo.science.kew.org, accessed 24 May 2023), while nomenclatural details (species authors and description year) were retrieved using IPNI (https://ipni.org, accessed 24 May 2023). All other data processing was performed in R (R Core Team 2022). Any data points falling in oceans were flagged using the *coordinateCleaner* R-package (Zizka *et al.* 2019). Points less than 15 km from the coastline were retained using custom R-scripts, while points outside that margin were removed (accounting for georeferencing uncertainty in coastal regions accidentally placing occurrences in the ocean). We generally report our results at the specimen level, thus separating duplicates of the same gathering (same collector and number), with exception of the expert curated database of Malagasy records. In that dataset, records refer to gatherings (i.e., one record might comprise multiple specimen sheets from multiple herbaria originating from the same gathering). For this reason, comparisons for Madagascar were made at the specimen level between the Herbonautes and GBIF datasets, and at the species level when considering the comparison between Herbonautes, GBIF and the expert dataset.

Specimens were projected onto a spatial raster with the resolution of 1 x 1 degrees (*c.* 110 × 110 km at the equator) using the WGS84 (ESPG:4326) coordinate reference system for global scale analyses, and 0.5 × 0.5° for regional analyses in Madagascar. For every grid-cell, collecting activities (collector names, number of specimens (only Herbonautes and GBIF) and number of species) were quantified. Differences in spatial coverage between the datasets were quantified using occurrence on the spatial grid at the respective resolutions. These resolutions balance computational efficiency with a relatively good level of detail at each respective scale. All scripts for analyses are available on github (https://github.com/SJRStreiff/Herbonautes_Annonaceae).

RESULTS

During the five Herbonautes 'episodes', 20 738 specimens were transcribed, georeferenced, databased and validated. Episode 1 focused on Madagascar, from July 2020 to November 2020 entering data for 2697 specimens in total. The second mission focused on Asia and was the one with the most specimens entered (8 308), lasting from December 2020 to end August 2021. Episode 3 focused on Oceania and was the smallest with 639 specimens entered between September and October 2021. The 2 353 specimens of the episode four focusing on the Americas lasted from October 2021 to December 2021. Finally, the project ended with Africa (Episode 5), the second largest episode, with 6739 specimens entered between June 2021 to February 2023. Thus, the transcription of the 20 738 specimens took two years and two months implicating 154 Herbonautes volunteers. However, participation was not equally distributed, with nine volunteers contributing over 90% of records transcribed. All metadata was retained and will be made publicly accessible (https://science.mnhn.fr/, or available from the first author on request).

Of the 20 738 records databased, 87 were marked as not usable (e.g., no label, label covered by specimen, or multiple specimens on a sheet). A further 3 541 specimens were not identified to species level, while an extra 188 records had a determination which could not readily be reconciled with the nomenclatural databases (predominantly invalidly published or unpublished manuscript names, e.g., *nomen in herb.*, *nomen dubium*, Turland *et al.* 2018, e.g., Fig. 1). Importantly, 9 404 specimens (*c.* 55 % of specimens after cleaning) had an additional expert identification. Note, however, in the case of multiple specimens for same gathering, a single determination slip is typically present on only one of the specimens. This leads to an underestimation of the total number of determined specimens in our Herbonautes dataset. After cleaning, 15 265 records were retained for spatial analyses out of the 20 738 original records. For the comparison dataset from GBIF, 75 441 records were retained after removing erroneous or missing coordinates and crossreferencing the taxonomy with POWO. In total, only 78 records were shared between the datasets, resulting from previous submissions of the herbarium P to the GBIF database.

These cleaned datasets contained 1 198 (Herbonautes) and 1 889 species (GBIF), respectively. In both datasets the number of specimens are unequally distributed across species; Coincidentally both for GBIF and the Herbonautes, 50% of the specimens belong to only 95 species. This equates to *c.* 5% and *c.* 8% for GBIF and Herbonautes respectively (Fig. 2). However, only 27 species are shared between these two subsets of 95 species.

In total, 159 species with 876 occurrences are contained in the Herbonautes dataset that had no records in GBIF. Of these 159 species, 55% can be considered endemics, with species only known to occur in a single "botanical country" (as defined by WGSRPD; https://www.tdwg.org/standards/wgsrpd/) according to POWO.

The Malagasy datasets contained 2 399 (expert), 1 902 (Herbonautes) and 3 501 (GBIF) records. The expert dataset contained most species (110, compared to 95 [P] and 90

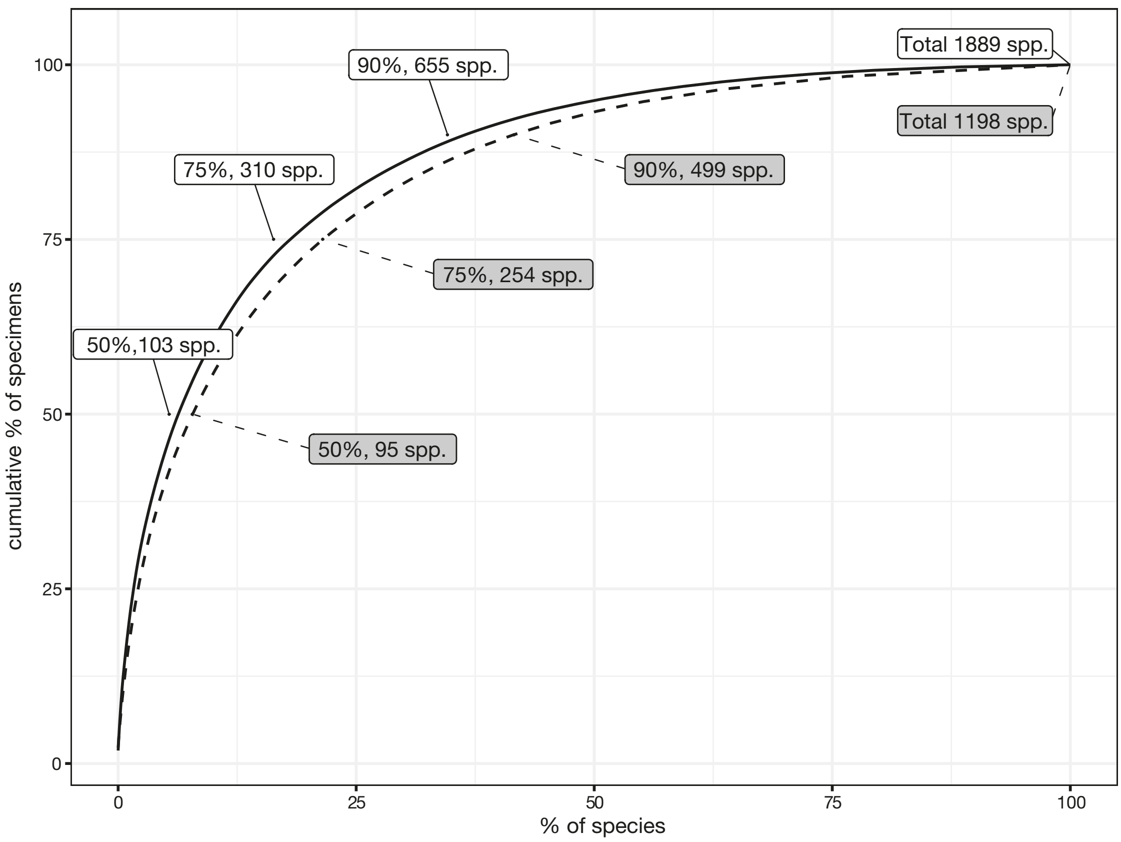


Fig. 2. — Cumulative percentage of specimens representing percentage of species, by dataset. The dashed line with grey boxes represents the Herbonautes dataset, continuous line with white boxes represent the GBIF dataset.

[GBIF]). The species missing in the Malagasy Herbonautes and GBIF subsets compared to the expert dataset are not equal: while species not found in the Herbonautes data are generally very recently described ones (median year of description 2013 [1852-2020]), the species missing in the GBIF dataset are more broadly distributed in time (median description year 1957 [1852-2020]).

From a spatial perspective, the Herbonautes database showed the highest number of species and specimens in tropical Southeast Asia as well as in central Africa and Madagascar (see Fig. 3 and Table 1). Specifically, the highest concentration of records per 1 x 1 degree square were located in western Cameroon, southern Viet Nam, coastal Côte d'Ivoire, northern Madagascar, central Democratic Republic of the Congo, northern Gabon and southwestern Central African Republic (Fig. 3A). Correspondingly, the three most collected countries were Madagascar, Viet Nam, Cameroon, making up a quarter of all records. Around half (49%) of all records were contained in the three previously mentioned countries plus Gabon, Malaysia, Côte d'Ivoire, French Guiana, Cambodia, Indonesia, the Republic of the Congo. These, together with the Philippines and New Caledonia also correspond to the countries visited by the highest number of different collectors (Fig. 3C). These patterns are also widely congruent with the number of species collected per grid cell (Fig. 3C). Furthermore, these regions generally correspond to areas where the Herbonautes dataset provides geographical coverage beyond previously existing GBIF data (Fig. 4). For the local comparisons in Madagascar, large differences in spatial coverage can be noted with both Herbonautes and the expert dataset having a similar coverage, whereas GBIF data is much more restricted (Fig. 5).

Table 1. — Species per dataset and continent.

Region	GBIF	Herbonautes	Herbonautes and not found in GBIF
Americas	700	248	7
Africa	268	189	17
Madagascar	79	79	10
Asia and Oceania	621	492	125

Collection efforts of Annonaceae in the Herbonautes dataset varied over time, with peaks in the second half of the 19*th* century, 1920s and the 1970s, and decreases – as expected - during the two World Wars. Towards the present, collection activities have once again decreased (Fig. 6). These patterns vary slightly from country to country, but remain generally comparable.

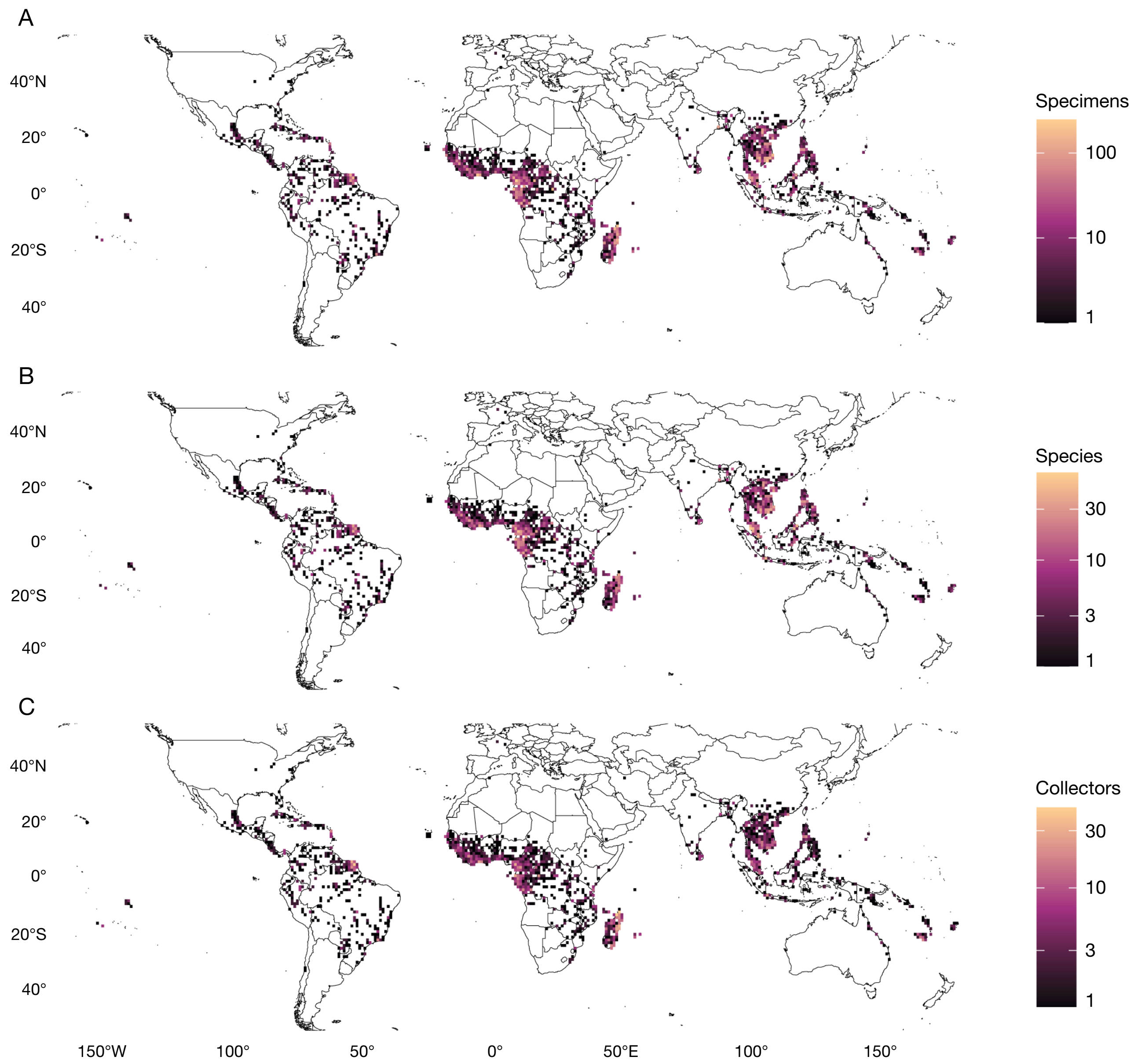


Fig. 3. — Spatial distribution and density of Annonaceae records transcribed by the 'Herbonautes' citizen science project, grouped by specimens (**A**, an individual specimen being defined as a single barcoded sheet), by species (**B**) and by botanical collectors (**C**), log scaled, grid resolution 1 × 1° (*c.* 110 × 110 km at the equator), equirectangular (EPSG 4326) projection.

DISCUSSION

Digitising herbarium specimens, especially the transcription of label metadata into databases, can pose significant challenges and be resource intensive. However, resulting datasets are of great value to understand biodiversity in many aspects. To quantify how the 'Herbonautes' citizen science transcription efforts led by the P herbarium allows for a better understanding of tropical diversity, we compared newly generated data for the angiosperm family Annonaceae with data previously available in GBIF. Overall, we show that the Herbonautes dataset led to an increase of 20% of digitally available georeferenced Annonaceae specimens and provided spatial data for 159 extra species when compared to GBIF (Table 1). Thus, to date, at least one record is openly and easily accessible for 1 938 out of 2 503 species (77.5%) of Annonaceae (Nge *et al.* in press).

The most important regions for which the Herbonautes dataset contributed data versus GBIF at the time of our study were located in South East Asia, Madagascar as well as parts of Central Africa (Fig. 4). These also represent a majority of countries formerly under colonial rule of France. Older specimens in the Herbonautes dataset originating from countries subject to colonisation by other European states are mostly duplicates: These duplicate specimens were usually collected in central herbaria and distributed through a network of specimen exchanges among other herbaria (Kaiser 2022). The major collecting regions highlighted were also visited by prominent collectors of the Paris herbarium, such as Eugène

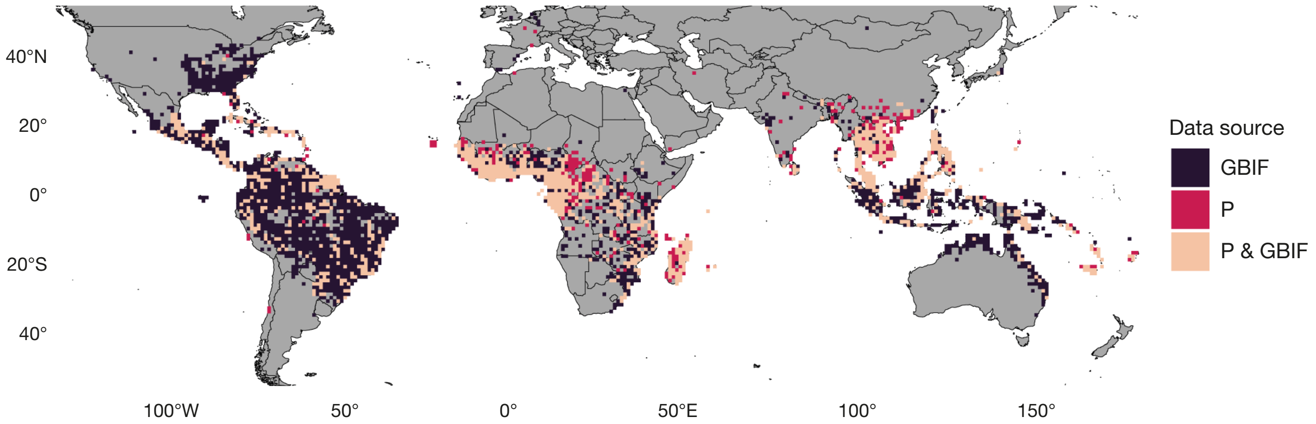


Fig. 4. — Global spatial coverage of datasets, showing regions covered by both Herbonautes (P) and GBIF data, just by GBIF and just by Herbonautes. Grid resolution 1 × 1° (*c.* 110 × 110 km at the equator), equirectangular (EPSG 4326) projection.

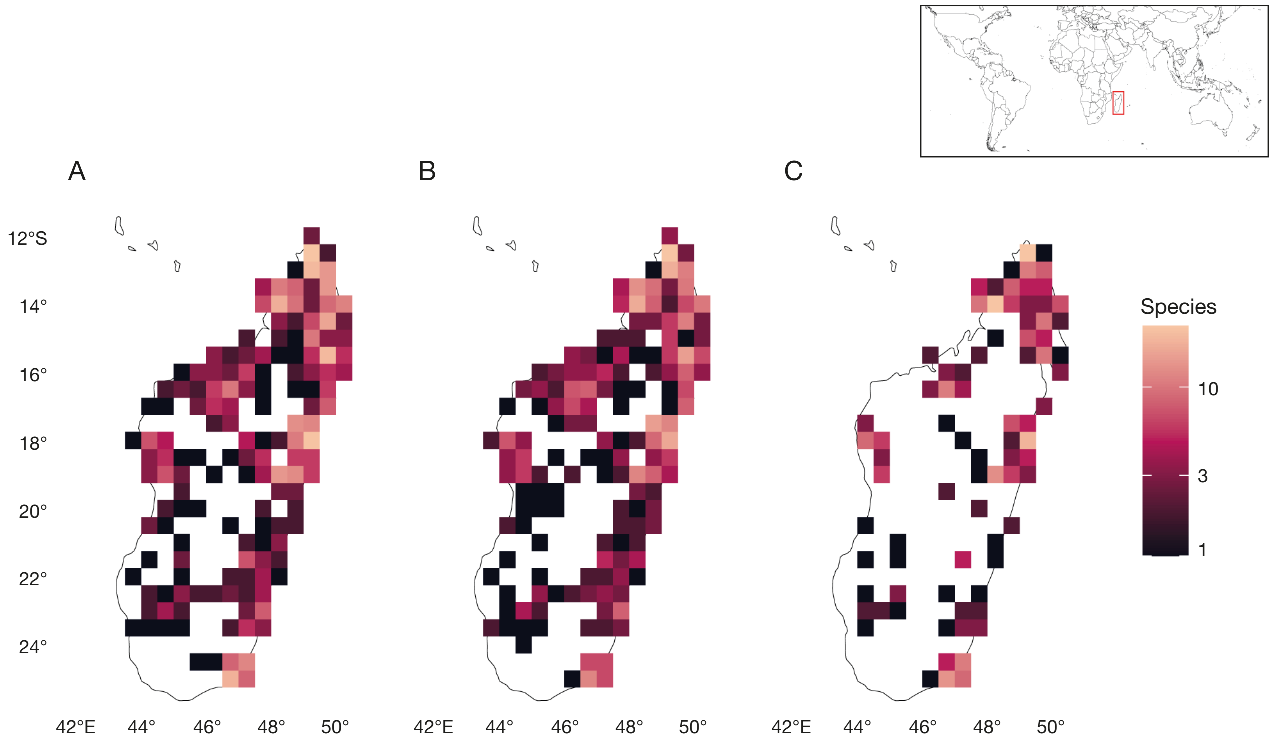


Fig. 5. — Spatial distribution of species richness in datasets for Madagascar at 0.5 × 0.5° grid resolution: **A**, curated expert dataset; **B**, herbonautes transcribed data; **C**, GBIF data. Equirectangular (EPSG 4326) projection.

Poilane (1888-1964; Burgos & Carré 2021), Louis Pierre (1833-1905; Leandri 1963), Auguste Chevalier (1873-1956; Bonneuil 1996) and René Capuron (1921-1971; Leroy 1972), to name a few.

More recently, collecting efforts are focused on those regions in which highly active taxonomic work is being performed. Such and similar variation in collecting efforts in time, and causes thereof, are starting to be addressed across collections and databases (Haripersaud *et al.* 2010; Zizka *et al.* 2021). For tropical Africa, in Benin (Akoègninou *et al.* 2006), Gabon (Sosef *et al.* 2006; Texier *et al.* 2022), and for Annonaceae specifically in Cameroun (Couvreur *et al.* 2022), the floras have been investigated thoroughly in recent years. This contributes to the data in these countries (among others) being generally better represented in GBIF, where also a small but important number of specimens were already transcribed for the respective projects at Paris and subsequently submitted. However only 5% of these records were georeferenced at that point in time. The other records were georeferenced during the respective Herbonautes episodes. Thus, the map (Fig. 4) does not completely reflect the impact of P specimens for Annonaceae diversity studies. Indeed, if we look at a country that has not been the focus of recent floristic studies for Annonaceae, such as Madagascar, we see that the Herbonautes dataset has a much wider spatial coverage when compared to GBIF (Fig. 6). Impor-

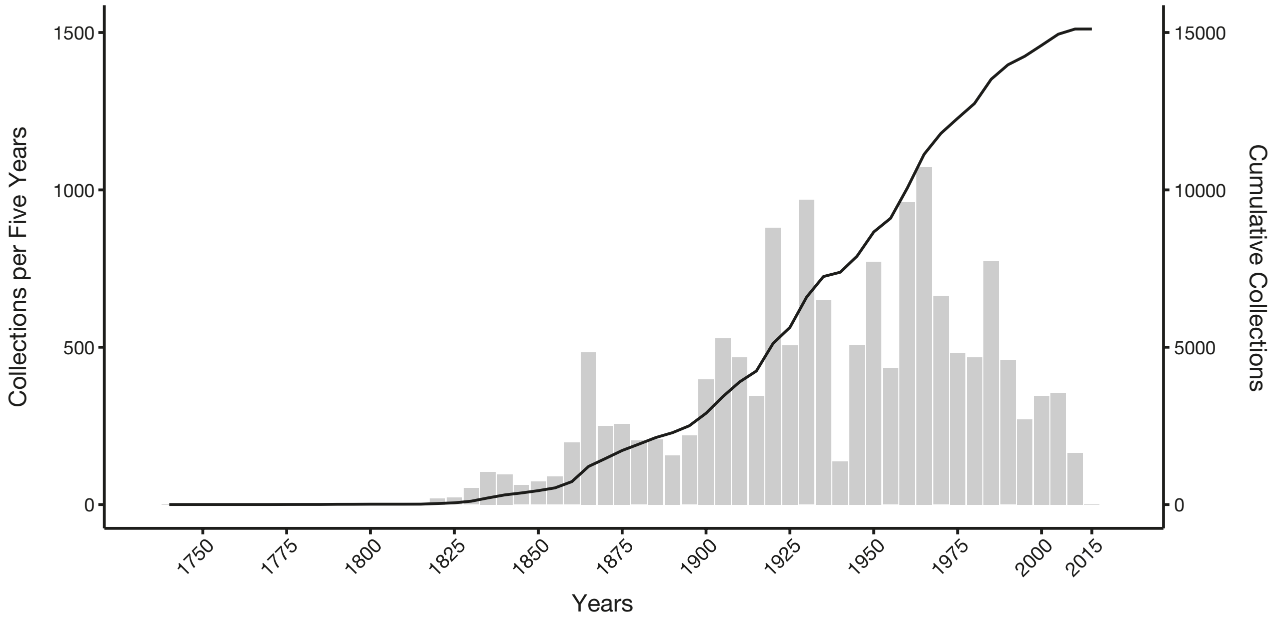


Fig. 6. — Temporal distribution of Annonaceae specimens collected in the Herbonautes dataset. The Histogram and left hand axis represent specimens collected per 5-year intervals. The right-hand axis and continuous line represent the cumulative specimens collected in total over the entire time period. The earliest Annonaceae collected and transcribed within the dataset is from 1740, a specimen of *Annona squamosa* L. collected in China by Pierre Nicolas le Chéron d'Incarville. The newest transcribed specimens are from 2015.

tantly, the Herbonautes data provided a significant portion of occurrence records for species of Madagascar (as seen by the relatively small differences between the Herbonautes and expert-curated datasets). Significant differences remain in spatial and taxonomic coverage when comparing these Malagasy datasets to GBIF. In particular, despite having a larger number of individual records, data from GBIF cover a significantly smaller geographic range than the other datasets. Species not present in the Herbonautes dataset, compared to the expert database, are recently new to science or placed in recently revised genera and are generally quite rare (e.g., Deroin & Gautier 2008; Gautier & Deroin 2013; Johnson & Murray 2020).

Nevertheless, the small differences above underline the value and importance of the specimens housed in the P herbarium, as such patterns may be even more prominent when considering families studied less frequently than Annonaceae. Here, important portions of diversity knowledge might remain in the storage cabinets and image servers of the P herbarium. Transcribing and making this data digitally available will undoubtedly contribute important information to systematic, ecological and conservation studies (Bebber *et al.* 2010). As in any natural history collection, collecting efforts vary across time (Fig. 6), but general trends show a recent slowdown. In addition, collections are increasingly also coordinated by more local herbaria. As these are associated with researchers and volunteers with in-depth knowledge of the local flora, the digitisation of these collections is just as important (Delves *et al.* 2024).

For the transcription of Annonaceae by 'Les Herbonautes' specifically, the main challenge presented by the up-front bulk imaging of specimens with transcription taking place later may be surprising: Keeping servers online and software running presented more problems on a project of the scale of 'Les Herbonautes' compared to questions or issues related to actual specimen label transcription. These problems tended to be either indecipherable handwriting or location names that could not be assigned to a present location, or that the specimen label was covered by the specimen on the actual specimen image. This affected only 78 specimens (<0.5%). Even with questions on handwriting or locations, crucially, the transcription of specimens by citizen scientists greatly frees up herbarium researchers and volunteers, especially when coordinated by a centralised structure or program. This centralised structure also facilitates and can help justify the informatic resources needed for the transcription of herbaria.

Transcribing the data stored in herbaria has potentially significant and far-reaching impacts. A significant number of undescribed species are estimated to be stored in herbaria (Bebber *et al.* 2010; Ondo *et al.* 2024). In the specific case of Annonaceae, previously documented patterns of species distributions noted a lack of available data (Erkens *et al.* 2023). Given that Annonaceae are an important part of tropical rain forest diversity globally (Turner 2001), and among the more extensively studied families in that biome, it is illustrative to demonstrate the difference these newly transcribed specimens have on global patterns of species richness: A significant number of Asian species were previously not digitally present, thus skewing previous estimates towards the neotropical forests. In extension, this has potentially important implications on past studies of global estimates of tropical plant diversity: Generally, tropical Asian tree species richness might be underestimated significantly (e.g., Antonelli *et al.* 2015; Kusumoto *et al.* 2023), in turn impacting large scale conservation assessments and conservation planning (Hoffmann *et al.* 2008; Pfab *et al.* 2011).

Leveraging platforms such as 'Les Herbonautes', numerous specimens can be transcribed accurately, reliably and relatively cost-effectively (the transcription effort for the 20 738 specimens being estimated as equivalent of 1.59 full-time equivalent (FTE) over the 'mission' period of approximately 24 months). Making such herbarium specimen data digitally accessible is a crucial step facilitating further studies into the ecology and evolution, for example in providing baseline data for species distribution modelling, in efforts to safeguard against extinction (Corlett 2023) and in conservation efforts such as Red Listing. Indeed, for most tropical plants, herbarium specimens are the basis of the assessments following IUCN criteria (Schatz 2002; Verspagen & Erkens 2023). While not replacing new collecting efforts in any way, digitising existing herbarium specimens is more cost-effective than collecting new specimens. Making these backlogs of specimens in herbaria worldwide available will have profound impacts on understanding past, present and future global biodiversity patterns along with biases inherent to collections (Meineke & Daru 2021; Corlett 2023; Davis 2023). This underlines that conservation managers and biodiversity policy-makers need to emphasise the funding of digitisation efforts in order to ensure proper conservation planning action.

CONCLUSION

Citizen science presents a powerful tool to reliably and accurately transcribe label information of herbarium specimen images. We illustrated this by the transcription of the more than 20 000 Annonaceae specimen images from the P herbarium. Transcribing records in such fashion also illustrates the gains that can be made by transcribing entire (sub-) collections, particularly in terms of species representation and geographic coverage. Moreover, the main challenges faced were not of the actual transcription nature, they were more commonly system related, emphasising the computational and informatic resources required for projects of this nature and size. Digitising, databasing and making the accumulated backlogs of herbarium specimens worldwide available digitally undoubtedly has profound impacts on our understanding of biodiversity in general, especially because herbaria are known to harbour numerous collections of yet to be described species. In addition, there is a real challenge of understanding and quantifying biases inherent with the collecting habits of botanists as well as curation habits of institutions. Finally, making data available digitally also represents a valuable resource for species distribution modelling and an effective pathway to improve conservation planning, especially IUCN Red Listing. These challenges will be better addressed with complete or near complete datasets, alleviating the uncertainty of existing but unavailable data.

Acknowledgements
We would like to acknowledge all the unnamed collectors who collected specimens. We would like to extend our gratitude to the 154 volunteers who as part of the "Les Herbonautes" project episodes painstakingly transcribed data from specimens. The curators and researchers who initiated, administrate and maintain the "Les Herbonautes" project and the herbarium are thanked for their valuable contributions. Margot Besseiche and Claude Patrick Millet provided valuable feedback on an early version of this paper. We are grateful to Drs Marc Sosef (Meise Botanic Garden Agency), Jerôme Munzinger (Institut de Recherche pour le Développement) and Thierry Deroin (Muséum national d'Histoire naturelle), their insightful comments and constructive feedback greatly improved the quality of this manuscript.
This project has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation program (grant agreement No. 865787, GLOBAL project to TLPC).
"Les Herbonautes" is a tool developed by the Récolnat infrastructure (ANR-11-INBS-0004).

REFERENCES

Akoègninou A., Burg W. J. van der & Maesen L. J. G. van der 2006. — *Flore analytique du Bénin*. Backhuys Publishers, 1034 p.

Albani Rocchetti G., Armstrong C. G., Abeli T., Orsenigo S., Jasper C., Joly S., Bruneau A., Zytaruk M. & Vamosi J. C. 2021. — Reversing extinction trends: new uses of (old) herbarium specimens to accelerate conservation action on threatened species. *New Phytologist* 230 (2): 433-450. https://doi.org/10.1111/nph.17133

Antonelli A., Zizka A., Silvestro D., Scharn R., Cascales-Miñana B. & Bacon C. D. 2015. — An engine for global plant diversity: highest evolutionary turnover and emigration in the American tropics. *Frontiers in Genetics* 6: 1-14. https://doi.org/10.3389/fgene.2015.00130

Bebber D. P., Carine M. A., Wood J. R. I., Wortley A. H., Harris D. J., Prance G. T., Davidse G., Paige J., Pennington T. D., Robson N. K. B. & Scotland R. W. 2010. — Herbaria are a major frontier for species discovery. *Proceedings of the National Academy of Sciences* 107 (51): 22169-22171. https://doi.org/10.1073/pnas.1011841108

Bonneuil C. 1996. — Auguste Chevalier, savant colonial : entre science et Empire, entre botanique et agronomie, *in* Waast R. (ed.), *Les sciences hors d'Occident au 20ème siècle* [*20th century sciences : beyond the metropolis*]; Vol. 2. *Les sciences coloniales : figures et institutions* [*Colonial sciences : researchers and institutions*]. ORSTOM, Paris: 15-35.

Burgos A. & Carré B. 2021. — Vie et œuvre d'Eugène Poilane (1888-1964). *Revue d'ethnoécologie* (20). https://doi.org/10.4000/ethnoecologie.7942

Carranza-Rojas J., Goeau H., Bonnet P., Mata-Montero E. & Joly A. 2017. — Going deeper in the automated identification of Herbarium specimens. *BMC Evolutionary Biology* 17 (181). https://doi.org/10.1186/s12862-017-1014-z

Chatrou L. W., Pirie M. D., Erkens R. H. J., Couvreur T. L. P., Neubig K. M., Abbott J. R., Mols J. B., Maas P. J. M., Saunders R. M. K. & Chase M. W. 2012. — A new subfamilial and tribal classification of the pantropical flowering plant family Annonaceae informed by molecular phylogenetics. *Botanical Journal of the Linnean Society* 169 (1): 5-40. https://doi.org/10.1111/j.1095-8339.2012.01235.x

Cheek M., Lughadha E. N., Kirk P., Lindon H., Carretero J., Looney B., Douglas B., Haelewaters D., Gaya E., Llewellyn T., Ainsworth A. M., Gafforov Y., Hyde K., Crous P., Hughes M., Walker B. E., Forzza R. C., Wong K. M. & Niskanen T. 2020. — New scientific discoveries: Plants and fungi. *Plants, People, Planet* 2 (5): 371-388. https://doi.org/10.1002/ppp3.10148.

Corlett R. T. 2023. — Achieving zero extinction for land plants. *Trends in Plant Science* 28 (8): 913-923. https://doi.org/10.1016/j.tplants.2023.03.019

Couvreur T. L. P., Pirie M. D., Chatrou L. W., Saunders R. M. K., Su Y. C. F., Richardson J. E. & Erkens R. H. J. 2011. — Early evolutionary history of the flowering plant family Annonaceae: steady diversification and boreotropical geodispersal. *Journal of Biogeography* 38 (4): 664-680. https://doi.org/10.1111/j.1365-2699.2010.02434.x

Couvreur T. L. P., Maas P. J. M., Meinke S., Johnson D. M. & Kessler P. J. A. 2012. — Keys to the genera of Annonaceae. *Botanical Journal of the Linnean Society* 169 (1): 74-83. https://doi.org/10.1111/j.1095-8339.2012.01230.x

Couvreur T. L. P., Helmstetter A. J., Koenen E. J. M., Bethune K., Brandão R. D., Little S. A., Sauquet H. & Erkens R. H. J. 2019. — Phylogenomics of the Major Tropical Plant Family Annonaceae Using Targeted Enrichment of Nuclear Genes. *Frontiers in Plant Science* 9: 1-15. https://doi.org/10.3389/fpls.2018.01941

Couvreur T. L. P., Dagallier L.-P. M. J., Crozier F., Ghogue J.-P., Hoekstra P. H., Kamdem N. G., Johnson D. M., Murray N. A. & Sonké B. 2022. — Flora of Cameroon. Vol 45: Annonaceae. *PhytoKeys* 207: 1-532. https://doi.org/10.3897/phytokeys.207.61432

Davis C. C. 2023. — The herbarium of the future. *Trends in Ecology & Evolution*. https://doi.org/10.1016/j.tree.2022.11.015

Delves J., Albán-Castillo J., Cano A., Fernández Aviles C., Gagnon E., Gonzáles P., Knapp S., León B., Marcelo-Peña J. L., Reynel C., Rojas Gonzáles R. del P., Rodríguez Rodríguez E. F., Särkinen T., Vásquez Martínez R. & Moonlight P. W. 2024. — Small and in-country herbaria are vital for accurate plant threat assessments: A case study from Peru. *Plants, People, Planet* 6 (1): 174-185. https://doi.org/10.1002/ppp3.10425

Deroin T. & Gautier L. 2008. — *Artabotrys darainensis* Deroin & L. Gaut. (Annonaceae), a new species from Madagascar. *Candollea* 63 (1): 93-99.

Dillen M., Groom Q., Chagnoux S., Güntsch A., Hardisty A., Haston E., Livermore L., Runnel V., Schulman L., Willemse L., Wu Z. & Phillips S. 2019. — A benchmark dataset of herbarium specimen images with label data. *Biodiversity Data Journal* 7 (e31817). https://doi.org/10.3897/BDJ.7.e31817

Drinkwater R., Cubey R. W. N. & Haston E. M. 2014. — The use of Optical Character Recognition (OCR) in the digitisation of herbarium specimen labels. *PhytoKeys* 38: 15-30. https://doi.org/10.3897/phytokeys.38.7168

Eloff J. N. 1999. — It is possible to use herbarium specimens to screen for antibacterial components in some plants. *Journal of Ethnopharmacology* 67 (3): 355-360. https://doi.org/10.1016/S0378-8741(99)00053-7

Engledow H., Smedt S. D., Groom Q., Bogaerts A., Stoffelen P., Sosef M. & Wambeke P. V. 2018. — Managing a Mass Digitization Project at Meise Botanic Garden: From Start to Finish. *Biodiversity Information Science and Standards* 2: e25912. https://doi.org/10.3897/biss.2.25912

Erkens R. H. J., Blanpain L. M. P., Carrascosa Jara I., Runge K., Verspagen N., Cosiaux A. & Couvreur T. L. P. 2023. — Spatial distribution of Annonaceae across biomes and anthromes: Knowledge gaps in spatial and ecological data. *Plants, People, Planet* 5: 520-535. https://doi.org/10.1002/ppp3.10321

Erkens R. H. J., Oosterhof J., Westra L. Y. T. & Maas P. J. M. 2017. — Revisions of *Ruizodendron* and *Pseudephedranthus* (Annonaceae) including a new species and an overview of most up-to-date revisions of Neotropical Annonaceae genera. *PhytoKeys* 86: 75-96. https://doi.org/10.3897/phytokeys.86.13773

Figueiredo E. & Smith G. F. 2010. — The colonial legacy in African plant taxonomy: biological types and why we need them. *South African Journal of Science* 106 (3): 1-4. https://doi.org/10.10520/EJC97032

Gallinat A. S., Russo L., Melaas E. K., Willis C. G. & Primack R. B. 2018. — Herbarium specimens show patterns of fruiting phenology in native and invasive plant species across New England. *American Journal of Botany* 105 (1): 31-41. https://doi.org/10.1002/ajb2.1005

Gautier L. & Deroin T. 2013. — *Uvaria lombardii* L. Gaut. & Deroin (Annonaceae), une nouvelle espèce endémique de Madagascar, aux inflorescences spectaculaires. *Candollea* 68 (2): 237-244. https://doi.org/10.15553/c2012v682a7

Goodwin Z. A., Harris D. J., Filer D., Wood J. R. I. & Scotland R. W. 2015. — Widespread mistaken identity in tropical plant collections. *Current Biology* 25 (22): R1066, R1067. https://doi.org/10.1016/j.cub.2015.10.002

Groom Q., Dillen M., Hardy H., Phillips S., Willemse L. & Wu Z. 2019. — Improved standardization of transcribed digital specimen data. *Database* 2019: baz129. https://doi.org/10.1093/database/baz129

Guo X., Tang C. C., Thomas D. C., Couvreur T. L. P. & Saunders R. M. K. 2017. — A mega-phylogeny of the Annonaceae: taxonomic placement of five enigmatic genera and support for a new tribe, Phoenicantheae. *Scientific Reports* 7 (1): 7323. https://doi.org/10.1038/s41598-017-07252-2

Haripersaud P., Steege H. ter, Granville J.-J. de, Chevillotte H. & Hoff M. 2010. — Species abundance, distribution and diversity in time and space after centuries of botanical collecting in the Guianas. *Taxon* 59 (2): 592-597. https://doi.org/10.1002/tax.592024

Heberling J. M., Prather L. A. & Tonsor S. J. 2019. — The Changing Uses of Herbarium Data in an Era of Global Change: An Overview Using Automated Content Analysis. *BioScience* 69 (10): 812-822. https://doi.org/10.1093/biosci/biz094

Hedrick B. P., Heberling J. M., Meineke E. K., Turner K. G., Grassa C. J., Park D. S., Kennedy J., Clarke J. A., Cook J. A., Blackburn D. C., Edwards S. V. & Davis C. C. 2020. — Digitization and the Future of Natural History Collections. *BioScience* 70 (3): 243-251. https://doi.org/10.1093/biosci/biz163

Hidalga A. N. de la, Rosin P. L., Sun X., Bogaerts A., Meeter N. D., Smedt S. D., Schijndel M. S. van, Wambeke P. V. & Groom Q. 2020. — Designing an Herbarium Digitisation Workflow with Built-In Image Quality Management. *Biodiversity Data Journal* 8: e47051. https://doi.org/10.3897/BDJ.8.e47051

Hoekstra P. H., Wieringa J. J., Maas P. J. M. & Chatrou L. W. 2021. — Revision of the African species of Monanthotaxis (Annonaceae). *Blumea – Biodiversity, Evolution and Biogeography of Plants* 66 (2): 107-221. https://doi.org/doi:10.3767/blumea.2021.66.02.01

Hoffmann M., Brooks T. M., Fonseca G. A. B. da, Gascon C., Hawkins A. F. A., James R. E., Langhammer P., Mittermeier R. A., Pilgrim J. D., Rodrigues A. S. L. & Silva J. M. C. 2008. — Conservation planning and the IUCN Red List. *Endangered Species Research* 6 (2): 113-125. https://doi.org/10.3354/esr00087

Johnson D. M. & Murray N. A. 2020. — A revision of *Xylopia* L. (Annonaceae): the species of Madagascar and the Mascarene islands. *Adansonia*, sér. 3, 42 (1): 1-88. https://doi.org/10.5252/adansonia2020v42a1. http://adansonia.com/42/1

Kaiser K. 2022. — Duplicate networks: the Berlin botanical institutions as a 'clearing house' for colonial plant material, 1891-1920. *The British Journal for the History of Science* 55: 279-296. https://doi.org/10.1017/S0007087422000139

Kirchhoff A., Bügel U., Santamaria E., Reimeier F., Röpert D., Tebbje A., Güntsch A., Chaves F., Steinke K.-H. & Berendsohn W. 2018. — Toward a service-based workflow for automated information extraction from herbarium specimens. *Database* 2018: bay103. https://doi.org/10.1093/database/bay103

Krishtalka L., Dalcin E., Ellis S., Ganglo J., Hosoya T., Nakae M., Owens I., Paul D., Pignal M. & Thiers B. 2016. — *Accelerating the Discovery of Biocollections Data*. Copenhagen, GBIF Secretariat. Available online at: http://www.gbif.org/resource/83022

Kusumoto B., Chao A., Eiserhardt W. L., Svenning J.-C., Shiono T. & Kubota Y. 2023. — Occurrence-based diversity estimation reveals macroecological and conservation knowledge gaps for global woody plants. *Science Advances* 9 (40): eadh9719. https://doi.org/10.1126/sciadv.adh9719

Leandri J. 1963. — Louis Pierre, botaniste de terrain et systématicien français (1833-1905). *Adansonia, nouvelle série,* 3 (2): 207-220. https://doi.org/10.5962/p.296340

Leroy J.-F. 1972. — René Capuron (1921-1971), fondateur de la Botanique forestière de Madagascar. *Adansonia, nouvelle série,* 12 (1): 13-38. https://doi.org/10.5962/p.296691

Le Bras G., Pignal M., Jeanson M. L., Muller S., Aupic C., Carré B., Flament G., Gaudeul M., Gonçalves C., Invernón V. R., Jabbour F., Lerat E., Lowry P. P., Offroy B., Pimparé E. P., Poncy O., Rouhan G. & Haevermans T. 2017. — The French Muséum national d'histoire naturelle vascular plant herbarium collection dataset. *Scientific Data* 4 (1): 170016. https://doi.org/10.1038/sdata.2017.16

Maas P. J. M. 1984. — The Annonaceae project. *Taxon* 33 (4): 800, 801.

Maas P., Westra L., Rainer H., Scharf U., Maas-van de Kamer H., Erkens R., Van Andel T. & Arias Guerrero S. 2023. — *Annonaceae*, *in* Flora of Guyana, Suriname and French Guiana. Muséum national d'Histoire naturelle, Paris; IRD, Marseille, 496 p. (Faune et Flore tropicales; 51)

McLauchlan K. K., Ferguson C. J., Wilson I. E., Ocheltree T. W. & Craine J. M. 2010. — Thirteen decades of foliar isotopes indicate declining nitrogen availability in central North American grasslands. *New Phytologist* 187 (4): 1135-1145. https://doi.org/10.1111/j.1469-8137.2010.03322.x

Meade C. V. & Parnell J. A. N. 2018. — A revised taxonomy for Uvaria (Annonaceae) in continental Asia. *Australian Systematic Botany* 31: 311-356. https://doi.org/doi:10.1071/SB17051

Meineke E. K. & Daru B. H. 2021. — Bias assessments to expand research harnessing biological collections. *Trends in Ecology & Evolution* 36 (12): 1071-1082. https://doi.org/10.1016/j.tree.2021.08.003

Nge F. J., Chaowasku T., Damthongdee A., Wiya C., Soulé V. R. C., Rodrigues-Vaz C., Bruy D., Mariac C., Chatrou L. W., Chen J., Choo L. M., Dagallier L.-P. M. J., Erkens R. H. J., Johnson D. M., Leeratiwong C., Lobão A. Q., Lopes J. C., Martínez-Velarde M. F., Munzinger J., Murray N. A., Neo W. L., Rakotoarinivo M., Ortiz-Rodriguez A. E., Sonké B., Thomas D. C., Wieringa J. J. & Couvreur T. L. P. (in press). — Complete genus-level phylogenomics and new subtribal classification of the pantropical plant family Annonaceae. *Taxon* (in press). https://doi.org/10.1002/tax.13260

Ondo I., Dhanjal-Adams K. L., Pironon S., Silvestro D., Deklerck V., Grace O. M., Monro A. K., Nicolson N., Walker B. & Antonelli A. 2023. — Plant diversity darkspots for global collection priorities preprint. *New Phytologist* 244 (2): 719-733. https://doi.org/10.1111/nph.20024

Park D. S., Feng X., Akiyama S., Ardiyani M., Avendaño N., Barina Z., Bärtschi B., Belgrano M., Betancur J., Bijmoer R., Bogaerts A., Cano A., Danihelka J., Garg A., Giblin D. E., Gogoi R., Guggisberg A., Hyvärinen M., James S. A., Sebola R. J., Katagiri T., Kennedy J. A., Komil T. S., Lee B., Lee S. M. L., Magri D., Marcucci R., Masinde S., Melnikov D., Mráz P., Mulenko W., Musili P., Mwachala G., Nelson B. E., Niezgoda C., Novoa Sepúlveda C., Orli S., Paton A., Payette S., Perkins K. D., Ponce M. J., Rainer H., Rasingam L., Rustiami H., Shiyan N. M., Bjorå C. S., Solomon J., Stauffer F., Sumadijaya A., Thiébaut M., Thiers B. M., Tsubota H., Vaughan A., Virtanen R., Whitfeld T. J. S., Zhang D., Zuloaga F. O. & Davis C. C. 2023a. — The colonial legacy of herbaria. *Nature Human Behaviour* 7 (7): 1059-1068. https://doi.org/10.1038/s41562-023-01616-7

Park D. S., Lyra G. M., Ellison A. M., Maruyama R. K. B., Dos Reis Torquato D., Asprino R. C., Cook B. I. & Davis C. C. 2023b. — Herbarium records provide reliable phenology estimates in the understudied tropics. *Journal of Ecology* 111 (2): 327-337. https://doi.org/10.1111/1365-2745.14047

Pfab M. F., Victor J. E. & Armstrong A. J. 2011. — Application of the IUCN Red Listing system to setting species targets for conservation planning purposes. *Biodiversity and Conservation* 20 (5): 1001-1012. https://doi.org/10.1007/s10531-011-0009-0

R Core Team 2022. — *R: A Language and Environment for Statistical Computing*. Vienna, Austria, R Foundation for Statistical Computing.

Ridge M. 2013. — From Tagging to Theorizing: Deepening Engagement with Cultural Heritage through Crowdsourcing. *Curator: The Museum Journal* 56 (4): 435-450. https://doi.org/10.1111/cura.12046

Rouhan G., Chagnoux S., Dennetiere B., Schäfer V. & Pignal M. 2016. — The herbonautes website: recruiting the general public to acquire the data from herbarium labels. *Botanists of the twenty-first century: roles, challenges and opportunities*. Paris, UNESCO.

Santos J., Cunha P. R. da & Sales F. 2020. — A strategy to digitise natural history collections with limited resources. *Biodiversity Data Journal* 8: e55959. https://doi.org/10.3897/BDJ.8.e55959

Silvertown J. 2009. — A new dawn for citizen science. *Trends in Ecology & Evolution* 24 (9): 467-471. https://doi.org/10.1016/j.tree.2009.03.017

Schatz G. E. 2002. — Taxonomy and Herbaria in Service of Plant Conservation: Lessons from Madagascar's Endemic Families. *Annals of the Missouri Botanical Garden* 89 (2): 145-152. https://doi.org/10.2307/3298559

Sosef M. S. M., Wieringa J. J., Jongkind C. C. H., Achoundong G., Issembe Y. A., Bedigian D., Berg R. G. van den, Breteler F. J., Cheek M., Degreef J., Faden R. B., Goldblatt D., Maesen L. J. G. van der, Banak L. N., Niangadouma R., Nzabi T., Nziengui B., Rogers Z. S., Stévart T., Valkenburg J. L. C. H. van, Walters H. M. J. & Wilde J. J. F. E. de 2006. — Check-list des plantes vasculaires du Gabon [*Checklist of Gabonese vascular plants*]. *Scripta Botanica Belgica* 35: 1-438. National Botanic Garden of Belgium, Meise.

Souza E. N. F. & Hawkins J. A. 2017. — Comparison of Herbarium Label Data and Published Medicinal Use: Herbaria as an Underutilized Source of Ethnobotanical Information. *Economic Botany* 71 (1): 1-12. https://doi.org/10.1007/s12231-017-9367-1

Sweeney P. W., Starly B., Morris P. J., Xu Y., Jones A., Radhakrishnan S., Grassa C. J. & Davis C. C. 2018. — Large-scale digitization of herbarium specimens: Development and usage of an automated, high-throughput conveyor system. *Taxon* 67 (1): 165-178. https://doi.org/10.12705/671.10

Texier N., Lachenaud O., Boupoya A., Bidault E., Fadeur G., Hardy O., Lowry P. P. II & Stévart T. 2022. — Characteristics and Determinants of Endemic Plant Taxa in the Gabonese Area of Endemism. *Annals of the Missouri Botanical Garden* 107: 1-20. https://doi.org/10.3417/2022694

Thiers B. (continuously updated). — *Index Herbariorum: A Global Directory of Public Herbaria and Associated Staff.* https://sweetgum.nybg.org/science/ih/, New York Botanical Garden's Virtual Herbarium.

THIERS B., TULIG M. C. & WATSON K. A. 2016. — Digitization of The New York Botanical Garden Herbarium. *Brittonia* 68 (3): 324-333. https://doi.org/10.1007/s12228-016-9423-7

TURLAND N. J., WIERSEMA J. H., BARRIE F. R., GREUTER W., HAWKSWORTH D. L., HERENDEEN P. S., KNAPP S., KUSBER W.-H., LI D.-Z., MARHOLD K., MAY T. W., MCNEILL J., MONRO A. M., PRADO J., PRICE M. J. & SMITH G. F. 2018. — *International Code of Nomenclature for algae, fungi, and plants (Shenzhen Code) adopted by the Nineteenth International Botanical Congress Shenzhen, China, July 2017*.Glashütten: Koeltz Botanical Books. (Regnum Vegetabile 159).

TURNER I. M. 2001. — *The Ecology of Trees in the Tropical Rain Forest*. Cambridge, Cambridge University Press. (Cambridge Tropical Biology Series). https://doi.org/10.1017/CBO9780511542206

VAN ANDEL T., VOS R., MICHELS E. & STEFANAKI A. 2022. — Sixteenth-century tomatoes in Europe: who saw them, what they looked like, and where they came from. *PeerJ* 10 (e12790). https://doi.org/10.7717/peerj.12790

VERSPAGEN N. & ERKENS R. H. J. 2023. — A method for making Red List assessments with herbarium data and distribution models for species-rich plant taxa: Lessons from the Neotropical genus Guatteria (Annonaceae). *Plants, People, Planet* 5 (4): 536-546. https://doi.org/10.1002/ppp3.10309

VOGEL ELY C., BORDIGNON S. A. DE L., TREVISAN R. & BOLDRINI I. I. 2017. — Implications of poor taxonomy in conservation. *Journal for Nature Conservation* 36: 10-13. https://doi.org/10.1016/j.jnc.2017.01.003

WEAVER W. N., RUHFEL B. R., LOUGH K. J. & SMITH S. A. 2023. — Herbarium specimen label transcription reimagined with large language models: Capabilities, productivity, and risks. *American Journal of Botany* 110 (12): e16256. https://doi.org/10.1002/ajb2.16256

WILKINSON M. D., DUMONTIER M., AALBERSBERG IJ.J., APPLETON G., AXTON M., BAAK A., BLOMBERG N., BOITEN J.-W., DA SILVA SANTOS L. B., BOURNE P. E., BOUWMAN J., BROOKES A. J., CLARK T., CROSAS M., DILLO I., DUMON O., EDMUNDS S., EVELO C. T., FINKERS R., GONZALEZ-BELTRAN A., GRAY A. J. G., GROTH P., GOBLE C., GRETHE J. S., HERINGA J., 'T HOEN P. A. C., HOOFT R., KUHN T., KOK R., KOK J., LUSHER S. J., MARTONE M. E., MONS A., PACKER A. L., PERSSON B., ROCCA-SERRA P., ROOS M., VAN SCHAIK R., SANSONE S.-A., SCHULTES E., SENGSTAG T., SLATER T., STRAWN G., SWERTZ M. A., THOMPSON M., VAN DER LEI J., VAN MULLIGEN E., VELTEROP J., WAAGMEESTER A., WITTENBURG P., WOLSTENCROFT K., ZHAO J. & MONS B. 2016. — The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data* 3 (1): 160018. https://doi.org/10.1038/sdata.2016.18

WILLIS C. G., ELLWOOD E. R., PRIMACK R. B., DAVIS C. C., PEARSON K. D., GALLINAT A. S., YOST J. M., NELSON G., MAZER S. J., ROSSINGTON N. L., SPARKS T. H. & SOLTIS P. S. 2017. — Old Plants, New Tricks: Phenological Research Using Herbarium Specimens. *Trends in Ecology & Evolution* 32 (7): 531-546. https://doi.org/10.1016/j.tree.2017.03.015

ZIZKA A., SILVESTRO D., ANDERMANN T., AZEVEDO J., RITTER C. D., EDLER D., FAROOQ H., HERDEAN A., ARIZA M., SCHARN R., SVANTESON S., WENGSTROM N., ZIZKA V. & ANTONELLI A. 2019. — CoordinateCleaner: standardized cleaning of occurrence records from biological collection databases. *Methods in Ecology and Evolution* (10): 744-751. https://doi.org/10.1111/2041-210X.13152

ZIZKA A., RYDÉN O., EDLER D., KLEIN J., PERRIGO A., SILVESTRO D., JAGERS S. C., LINDBERG S. I. & ANTONELLI A. 2021. — Bio-Dem, a tool to explore the relationship between biodiversity data availability and socio-political conditions in time and space. *Journal of Biogeography* 48 (11): 2715-2726. https://doi.org/10.1111/jbi.14256

Submitted on 23 January 2024;
accepted on 7 May 2024;
published on 4 November 2024.