RESEARCH ARTICLE

The effectiveness and limitations of digital images for taxonomic research

Aireen Phang,[1,2,3] Hannah Atkins[1] & Peter Wilkie[1]

1 *Royal Botanic Garden Edinburgh, 20A Inverleith Row, Edinburgh, EH3 5LR, United Kingdom*
2 *Singapore Botanic Gardens, National Parks Board, 1 Cluny Road, 259569 Singapore*
3 *School of Biological Sciences, University of Aberdeen, Aberdeen AB24 3UU, United Kingdom*
Addresses for correspondence: *Aireen Phang, aphang@rbge.org.uk; Hannah Atkins, hatkins@rbge.org.uk; Peter Wilkie, pwilkie@rbge.org.uk*

DOI https://doi.org/10.1002/tax.12767

Abstract In response to the worldwide coronavirus outbreak, which effectively shut down fieldwork, laboratory and herbarium-based studies, an evaluation was made of the effectiveness and limitations of undertaking a virtual taxonomic study using only online herbarium specimen resources related to the genus *Madhuca* (Sapotaceae) for the *Flora of Singapore*. The study demonstrated the immense value of digital images to basic taxonomic research but also found that diagnostic micro-morphological characters, often critical in defining species boundaries, cannot be seen in many digital images, even at high resolution. Several recommendations are made on how to maximise the utility of online herbarium specimen images to help facilitate future taxonomic research, though it is clear that physical access to herbarium specimens remains essential.

Keywords digitisation; flora account; herbarium specimens; *Madhuca*; morphology; Sapotaceae

Supporting Information may be found online in the Supporting Information section at the end of the article.

INTRODUCTION

The worldwide coronavirus (COVID-19) pandemic has helped accelerate the move to online working across a wide range of sectors, with taxonomic research being no exception. Using findings from research on the genus *Madhuca* J.F. Gmel. undertaken during the U.K. and Singapore lockdowns of 2020, this study investigates the limits of what can be achieved with the wealth of herbarium specimen images now available online and in so doing, indicate where future efforts need to be targeted to maximise their utility.

Large amounts of resources are being invested by herbaria around the world to digitise their specimens and make these available online. Initiatives such as those by Naturalis (Netherlands), Digitarium (Finland), and the Muséum National d'Histoire Naturelle (France) have resulted in their herbaria (4 million, 5 million and 5.4 million herbarium specimens, respectively) being successfully digitised in just a few years. Other herbaria are involved in national or international networks to digitise their collections, such as the Atlas of Living Australia (ALA, 2021) with over 8 million specimen records across Australia and New Zealand (Belbin & al., 2021), the Distributed System of Scientific Collections (DiSSCo, 2021) in Europe (currently more than 130 institutions from 21 countries), and the Integrated Digitized Biocollections (iDigBio, 2021) in the U.S.A., which so far has made more than 20 million digital specimen records available (Le Bras & al., 2017; Borsch & al., 2020).

The JSTOR Global Plants database (https://plants.jstor.org/), a major digital resource, has made more than two million type specimen images, arguably the most important kind of herbarium specimens, available from scores of herbaria across the world. Websites focused on particular taxonomic groups have also provided access to digital specimen information, such as Solanaceae Source (2021) and the Royal Botanic Garden Edinburgh's searchable databases for Begoniaceae, Sapotaceae and Zingiberaceae (RBGE, 2021). Other websites with a more geographical focus, e.g., the Flora do Brazil (Reflora – Virtual Herbarium, 2021), *species*Link (CRIA, 2021), the Andes to Amazon Biodiversity Information System (Atrium, 2021) and the Flora of Nepal (2014–) are also valuable sources of specimen images.

The term digitisation can be applied to several different activities when referring to herbarium specimens. In some cases, it refers only to the production of digital images of specimens, in others it is only the transcription of the data from herbarium labels onto a database; however, it is mostly used to refer to the production of a digital image of a specimen

Article history: Received: 12 Nov 2021 | returned for (first) revision: 7 Jan 2022 | (last) revision received: 8 Mar 2022 | accepted: 25 Apr 2022 | published online: 4 Jun 2022 | **Associate Editor:** Maria Vorontsova |
TAXON published by John Wiley & Sons Ltd on behalf of International Association for Plant Taxonomy.

and the transcription of data from the specimen into a database (although the amount of data entered from the specimen label often varies greatly).

Image technology has been used for almost half a century to supplement access to actual specimens, from images on microfiche of the "Wallich Herbarium" (IDC [undated]) (Fig. 1A) and Herbarium Willdenow (Hiepko, 1972) to photographs printed and mounted onto a herbarium sheet and physically placed in the herbarium such as those of the Beccari Herbarium produced by the Museo di Storia Naturale di Firenze in Italy. (Herb. Univ. Florentinae, Fig. 1B). These were all responses to the need of taxonomists to see and examine specimens not easily accessible to them, but still required for taxonomic study. At the time, these images served as a valuable research tool, but with time they have become redundant as new technologies have emerged. Today, most digital images are of higher quality and are usually made freely available online so that anyone with a computer and access to the World Wide Web can view them, e.g., the Global Biodiversity Information Facility (GBIF, 2021), which captures specimen images from several major institutions with digital records.

The highest-quality specimen images available online tend to be from larger institutes (in this paper referred to as "institutional images") and many follow the JSTOR Global Plants Initiative guidelines for the digitisation of herbarium specimens (JSTOR, 2018a,b). These recommend 600 pixels/dots per inch (dpi) as the ideal scanning resolution (although the online image is usually smaller; Nieva de la Hidalga & al., 2020), as well as the addition of scale bars and colour charts to facilitate analysis and measurement. Many institutes have also inserted barcodes and provided access to metadata. Other specimen images available online vary in quality and are often images taken during very short visits to collections by researchers (in this paper referred to as "researcher images"). These are frequently made to facilitate the databasing of specimen information away from the institution being visited, to allow species distributions to be established and to aid the confirmation of species identifications rather than for detailed taxonomic study.

By including smaller herbaria that may not yet have the ability or resources to generate standard digital images from their specimens (Delves, 2021), researcher databases often

Fig. 1. A, An image from microfiche of the herbarium of the East India Company ("The Wallich Herbarium") at Kew; **B**, A black and white photograph of a Beccari Herbarium specimen from the Museo di Storia Naturale di Firenze mounted and incorporated into the E herbarium (barcode E00632145).

can capture a greater number of specimen records beyond the information held in large institutions. For instance, the Royal Botanic Garden Edinburgh's Sapotaceae Resource Centre (SRC) (Wilkie & al., 2008–) contains significantly more herbarium specimen records than the Global Biodiversity Information Facility (GBIF) for many taxa in that family (Fig. 2 for a comparison of some *Madhuca* species). Both researcher and institutional databases, which often contain field images linked to specimens and related metadata, help to further the "extended specimen concept" (Lendemer & al., 2020), which brings together all physical and digital data associated with a single specimen, including sample preparations, habit photographs, genomic and trait data.

Many papers covering digitisation have focused on process such as digitisation workflows and equipment (Blagoderov & al., 2012; Holovachov & al., 2014; Takano & al., 2019; Borsch & al., 2020; Nieva de la Hidalga & al., 2020; Davis & al., 2021), standards and best practice (Häuser & al., 2005; Baskauf & Kirchoff, 2008), novel ways of specimen data input such as Optical Character Recognition (OCR) (Drinkwater & al., 2014) and crowdsourcing (Zhou & al., 2018; King & al., 2019). Studies on the use of these images have concentrated on species identification or specific extractable traits, usually related to leaf morphology (Corney & al., 2012; Carranza-Rojas & al., 2017; Borges & al., 2020; Pryer & al., 2020).

Borges & al. (2020) noted that accurate morphological data, particularly measurable values, can be obtained from professionally scanned 2D specimen images. The focus of this paper is an assessment of the utility of digital images to taxonomists, arguably the key users of herbarium specimens, in delivering taxonomic accounts, in this case an account of *Madhuca* for the *Flora of Singapore* (2019–). In doing so it will examine the limits of using digital images in producing a taxonomic account and suggest potential improvements to address these. It will also make reference to the applicability to other plant groups.

Tropical forests contain at least 40,000 tree species (Slik & al., 2015), in comparison to around 124 in temperate European regions, and there is a significant knowledge gap in estimating accurate tropical tree species and their biodiversity codependencies. LaFrankie (1996) found 814 species in just a 50 ha sample of a lowland Malayan rainforest, and varying levels of species representation, with Sapotaceae trees (including *Madhuca*) in low abundance. Species in low densities are systematically undercollected (Middleton & al., 2019), which exacerbates the critical lack of taxonomic information necessary to preserve biodiversity. This underlines the need to accelerate the completion of tropical flora projects, yet progress to document floras in Southeast Asia has been relatively slow (Middleton & al., 2019), and the lack of family treatments in projects such as the *Flora of Peninsular Malaysia* and *Flora Malesiana* has downstream implications on conservation and land-use planning. As such, there is a need to accelerate the production of Southeast Asian Floras using every resource possible, whether digital or physical.

■ MATERIALS AND METHODS

A list of *Madhuca* species names recorded as occurring in Singapore and Johore was compiled from the literature

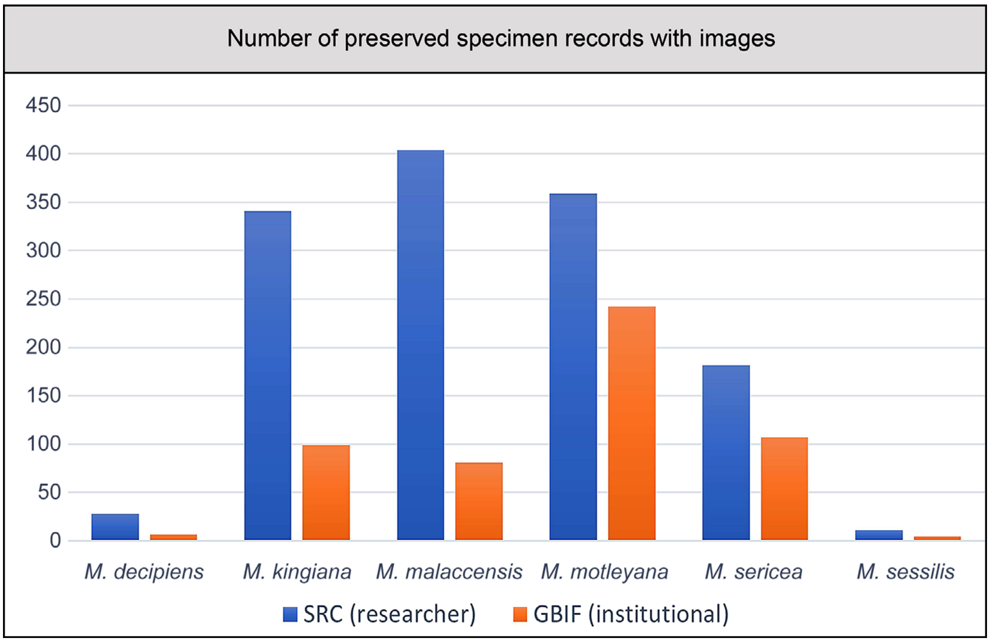

Fig. 2. Number of herbarium specimens with images in the Sapotaceae Resource Centre (SRC) compared to the Global Biodiversity Information Facility (GBIF), for selected species of *Madhuca* (*M. decipiens* J.Sinclair, *M. kingiana* (Brace ex King & Gamble) H.J.Lam, *M. malaccensis* (C.B.Clarke) H.J.Lam, *M. motleyana* (de Vriese) J.F.Macbr., *M. sericea* (Miq.) H.J.Lam, and *M. sessilis* (King & Gamble) Baehni).

(Assem, 1953; Royen, 1960; Ng, 1972; Turner, 1995; Chong & al., 2009) and online sources such as the World Checklist of Selected Plant Families (WCSP, 2021), the International Plant Names Index (IPNI, 2021) and the World Flora Online (WFO, 2021). Johore, as the southernmost state of Peninsular Malaysia closest to Singapore, was selected as a basis to compare affinities of *Madhuca* species found in Singapore.

Digital images of *Madhuca* specimens collected from Singapore and Johore were viewed from the following herbaria: E, K, KEP, L, SAN, SAR, SING (herbarium codes follow Thiers, 2021). These images were accessed via a range of sources, including the JSTOR Global Plants database (mainly for type specimens), online databases, and provided by staff from individual herbaria. The SRC was a major source of images, containing over 57,000 Sapotaceae records from 210 herbaria, and 21,000 images of both herbarium specimens and field photographs from around the world. In this study, if there were both institutional and researcher images of the same specimen, only the institutional image was used in the analysis.

Scanned specimen images of *Madhuca* species collected in Singapore and held in the herbarium of the Singapore Botanic Gardens (SING) were provided directly by herbarium staff. Specimen images held at the herbarium of the National University of Singapore (SINU) were not available due to institutional coronavirus-related closures.

With partial lifting of coronavirus-related restrictions, some physical access to the herbarium collections in SING was permitted and allowed the examination of physical specimens using a stereo dissecting microscope (Olympus SZX7) and light attachments. This allowed a comparison of taxonomic utility between physical specimens and digital images.

Institutional and researcher images were compared with information observed from physical examination of specimens and assessed for their utility in capturing important taxonomic detail. Floristic and taxonomic accounts of Sapotaceae were consulted to compile a list of characters used to describe and diagnose taxa (Table 1**)**. In this study, macro-morphological characters refer to those visible with the naked eye from the specimen or image, without additional optical aids. Micro-morphological characters are those visible only with a 10× hand lens or microscope, or discernible through touch (e.g., texture).

Based on these characters, the utility of a specimen image was rated as high, medium, or low. The availability of a scale bar or measuring tool was an important component of this assessment. Those images rated high allowed capture of at least 5 qualitative and 3 quantitative macro-morphological characters (NB. a single character considered here consists of all subcategories under each character per row in Table 1). Medium-ranked images allowed the capture of at least 5 qualitative macro-morphological characters. Low-ranked images allowed the capture of 4 qualitative macro-morphological characters, the minimum number found to distinguish discrete species.

Images from initial online resources were sorted virtually into taxon groups; this was replicated when access was obtained to actual specimens in the SING herbarium.

■ RESULTS

A total of 219 specimen images from various herbaria and online sources were examined: 95 were collected from Singapore and 124 from Johore (Table 2; suppl. Tables S1, S2). Of these images, 119 were of fertile specimens (those with flowers or fruits) with the rest of the specimens being sterile. Out of the 95 from Singapore, 11 specimens initially had no images and were available only after physical access to the SING herbarium and a request made to have them digitised. There were 114 specimens collected from Singapore and Johore examined in person.

All online herbarium specimen catalogues used allowed for straightforward searches by species name and location; however, only L, JSTOR and the SRC allowed for more narrowly defined searches by minor locality. For E and K herbaria, location searches were limited to country (sometimes only the larger herbarium filing region, i.e., a region considered by the herbarium to be floristically distinct, and may include several countries), hence specimens from the whole of Malaysia or herbarium filing region had to be sifted through to locate the specimens from the state of Johore. Images of specimens could be saved or downloaded from all databases; however, only E allowed for image download in a range of pixel sizes. The K and SING databases provided limited specimen information for the *Madhuca* species under study, with the former consisting mainly of data and images of type specimens, and the latter with occurrence records but no downloadable images. Institutions such as K do provide higher-quality images upon request; however, no applications were made in this study due to either the images already being the highest possible quality available, or because no specimen record was found in the online catalogue to provide a basis for the request.

The images in this study varied widely in quality. All images of *Madhuca* specimens collected in Singapore and held at SING were provided directly by the institute (as they were not yet available online); these had been scanned and macro-measurements could be taken using the appended scale bars. Other sources of institutional images were those downloadable directly from JSTOR, E, K and L online catalogues. Around half (116 out of 219) of all specimen images seen were assessed as high-quality scans, allowing detailed views of macro-morphological characters using zoom functions. The researcher images were not available to download directly from institutional websites (i.e., from KEP, SAN, SAR and SING) but had been made available by researchers and deposited in the SRC using a standard digital camera. As such, the quality of these images was more uneven, with most images not having a scale bar or colour chart, and in many cases, the capsule had not been opened to reveal its contents (Fig. 3A).

Ability to discern morphological characters. — There were clear limitations in observing characters from both institutional images and researcher images, and as such the characters initially prioritised as diagnostic were macro-morphological: e.g., petioles, leaves, shapes of leaf blades, midribs and venation details. While inflorescence sizes and some larger fertile

Table 1. Qualitative and quantitative characters required for a taxonomic description of *Madhuca* species found in Singapore

Qualitative and macro-morphological characters (scale bar not required)	Quantitative macro-morphological characters (scale bar required)	Micro-morphological characters, including textural
Leaf arrangement		
Blade - shape - apex - base	Blade - size	Blade - thickness and texture - indumentum above/below
		Midrib indentation/protrusion above/below
Petiole - shape	Petiole - length	Petiole - indumentum
Secondary veins - number - angle from midrib - type		Vein indentation/protrusion above/below
		Intersecondary veins - presence - type
		Tertiary veins - type
Stipules - presence - shape	Stipules - size	Stipules - indumentum
Inflorescence - number of flowers - placement		
	Pedicel - length	Pedicel - indumentum
		Sepals - number - shape - size - indumentum - margin
Corolla - colour		Corolla - length - indumentum
		Corolla tube - length
		Corolla lobes - number - shape - length - margin
		Stamens - number - presence/number of stamen rows
		Anthers - shape - length - indumentum - type of connective appendage
		Filament - length

(Continues)

Table 1. Continued.

Qualitative and macro-morphological characters (scale bar not required)	Quantitative macro-morphological characters (scale bar required)	Micro-morphological characters, including textural
		Ovary - number of locules - indumentum
Style - insertion/exsertion		Style - length - indumentum
Fruit - type - shape	Fruit - size - size of sepals - size of style	Fruit - number of seeds - lindumentum
Seed - shape	Seed - size	Seed - surface of testa - size/shape of seed scar

Table 2. Summary of number of specimen images studied.

Species	Collected from	Total images	Institutional images	Researcher images	Herbarium codes						
					E	K	KEP	L	SAN	SAR	SING
Madhuca decipiens	Singapore	17	14	3	2	4	–	1	–	–	10
	Johore	15	2	13	–	1	7	2	–	1	4
Madhuca kingiana	Singapore	49	49	0	2	–	–	1	–	–	46
	Johore	61	13	48	1	9	19	12	1	6	13
Madhuca malaccensis	Singapore	8	8	0	–	–	–	–	–	–	8
	Johore	13	4	9	–	1	3	4	1	–	4
Madhuca motleyana	Singapore	3	3	0	–	–	–	–	–	–	3
	Johore	20	3	17	1	2	7	2	–	–	8
Madhuca sericea	Singapore	9	7	2	–	1	1	–	–	–	7
	Johore	15	4	11	1	2	6	3	–	–	3
Madhuca sessilis	Singapore	2	2	0	–	1	–	–	–	–	1
	Johore	0	0	0	–	–	–	–	–	–	–
Incertae sedis	Singapore	7	7	0	–	–	–	–	–	–	7

characters could be seen and measured from images, without access to the physical specimens, micro-morphological attributes, including sepals, corollas, stamens, styles, margins, indumentum, fruit or seed details could not be discerned or described, even when small fertile parts or dissections contained within capsules were exposed (Fig. 3B: *Sinclair 39640*: capsule contents shown).

Of the characters used to produce the taxonomic account of *Madhuca* of Singapore, less than half of the characters could be measured and described from using images alone (no matter of what quality) and required access to physical specimens. Those characters that could be used, the macro-morphological characters, included: leaf phyllotaxy, blade shape and size, number of secondary veins, stipule shape and size, inflorescence placement and size. Micro-morphological characters could not be confidently described or measured from any of the images (institutional or researcher).

Of the 219 specimen images examined, 125 (comprising 103 researcher images and 22 institutional images) had macro-morphological characters hidden from view due to the low resolution of the image, the way the specimen had been mounted onto the herbarium sheet or had portions placed in an unopened capsule on the sheet. This included secondary venation characters that were difficult to distinguish due to poor lighting, partially concealed flowers on inflorescences, or seeds placed in capsules.

For 26 institutional images, the venation types initially recorded from the specimen image were corrected after physical examination of specimens in the herbarium. This was largely due to better lighting of the specimen, which highlighted details such as festooned brochidodromous or intramarginal vein characters.

Institutional images, which in all cases had scale bars, allowed measurements of macro-morphological characters and some subcategories such as outer sepal and pedicel lengths as well as seed and fruit size. Images on JSTOR could be easily measured with the built-in ruler functionality on the website, which allows automatic measurement of length or angles between two selected points. For other institutional images not from JSTOR, a physical protractor or ruler had to be tallied against the scale bar on the image, and then applied to the specimen features. Measurements could not be accurately taken from the researcher images in the SRC as no scale bars were included; however, the records were useful for capturing information on specimen labels (such as collector notes on habit or colours), as well as confirming some identifications and qualitative characters such as leaf arrangement, blade shape, venation patterns and relative inflorescence size. Image size and resolution had little impact on utility – many researcher images were of equivalent size and resolution to institutional ones (i.e., 300 dpi), and were only found to be of lower utility due to the lack of a scale bar. A summary of specimens examined is shown in Table 2. More details on each specimen can be found in suppl. Tables S1 and S2.

Image taxonomic utility rating. — Institutional images (comprising 53% of the total) received a high taxonomic utility rating mainly due to the ability to measure macro-morphological characters with the appended scale bar, whereas researcher images ranged from low to medium, dependent on the quality of the image to discern characters such as leaf venation or some fertile aspects. Flowering or fruiting specimens were often rated higher as there were additional morphological characters available. (Fig. 4A)

While four qualitative characters were the minimum required to distinguish the specimen images into morphologically discrete groups, this did not mean that the same characters were sufficient for species identification. Certain diagnostic characters are more distinctive for some species than for others, so a fixed set of characters could not be used to identify all species (Fig. 4B). As has been noted in several accounts of the Sapotaceae family, characters unique to a genus are rare (Pennington, 1991), hence species identifications require a suite of characters rather than a single variable one.

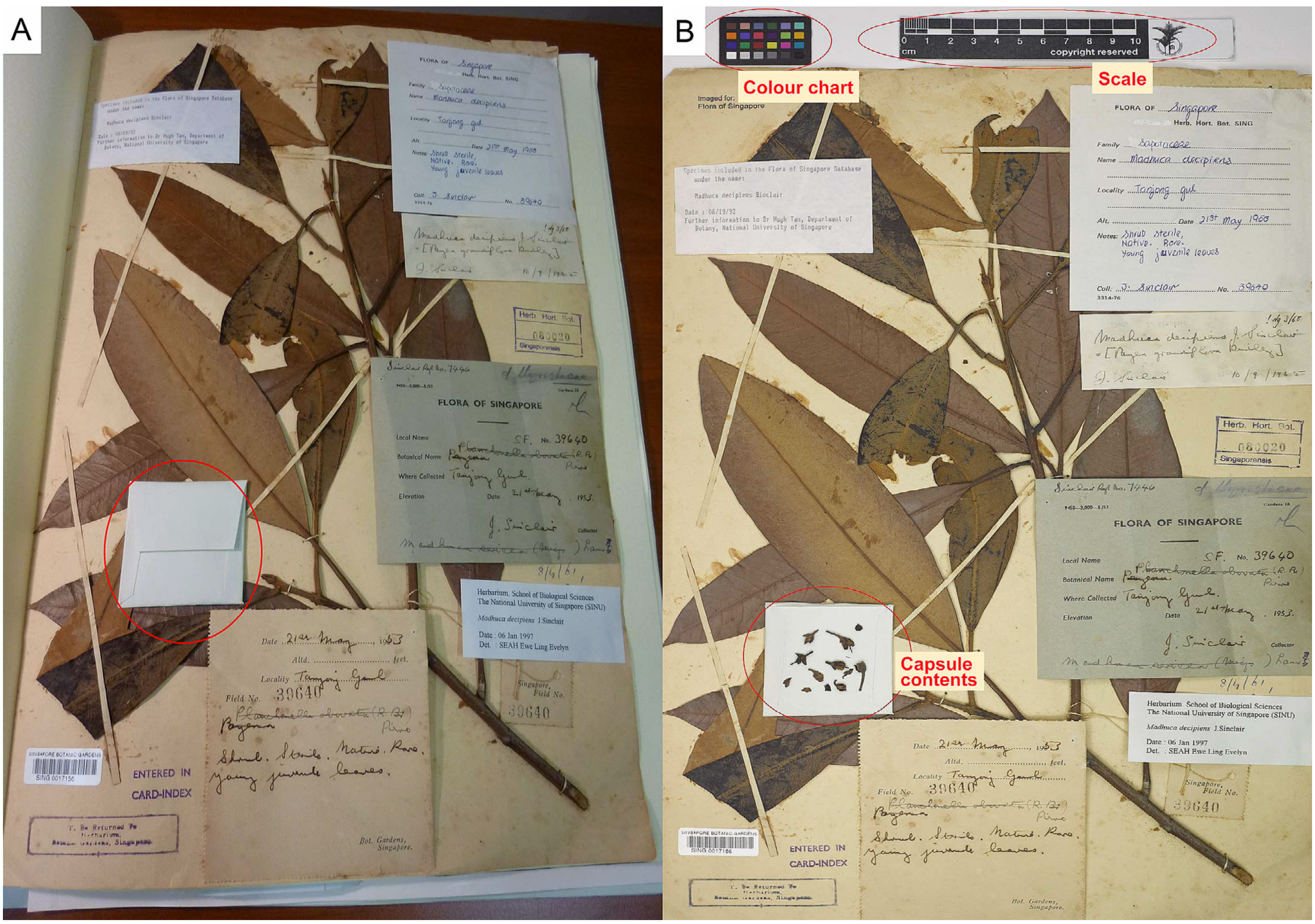

Fig. 3. Specimen images of the same collection (*Sinclair 39640*, SING) obtained from: **A**, Sapotaceae Resource Centre (SRC), with capsule unopened (circled in red); **B**, Professional scanning equipment at SING, with colour chart, scale bar and capsule contents laid out.

As many Sapotaceae genera have similar leaf arrangements and blade shapes (e.g., six genera, including *Madhuca* in Pennington's tribe Isonandreae, have spiral phyllotaxy), the reference for species identification began from images of the type specimen usually containing fertile material, and then cross-checked against the species description in the protologues. From features seen and described from the type specimen, specimen images could then be assessed for conspecificity.

Only two species, *Madhuca kingiana* and *M. sessilis*, could be positively identified with four qualitative characters from images alone, due to the distinctive diagnosis of secondary venation in the former, and sessile leaves in the latter. With a scale bar, the specimen images could be further distinguished based on leaf size, which then allowed *M. motleyana* to be identified with two additional quantitative characters (for blade size and petiole length) due to its mature leaves being smaller than the rest.

In all cases, species descriptions based on images alone (both institutional and researcher images) could incorporate macro-morphological characters (mostly vegetative), and less than half of the images could be confidently assigned to species (Fig. 4B).

For some specimen images, including institutional ones, indumentum and venation details could not be positively identified without access to the physical specimen and the use of a microscope (Fig. 5) with light attachment. Only after physical examination of actual herbarium specimens could species identifications be confirmed, and micro-morphological characters, especially floral, fruit and seed details, be incorporated into the taxonomic descriptions for the six species of *Madhuca* occurring in Singapore. There were, however, a few specimens with limited or low-quality material, hence identification was not possible even with direct examination (Fig. 4B).

Physical access to the collections in SING discovered three recent flowering collections, two of *Madhuca decipiens* (*Ng SING 2019-245* and *Lua & al. SING 2019-385*, April 2019), and one of *M. kingiana* (*Ngo SING 2019-207*, March 2019) that had not been included in the initial set of digitised images as they had been freshly mounted. These were particularly useful for studying internal flower morphology (Fig. 6A).

Although not directly related to the quality of images, it is relevant to note that fruit and seed measurements were challenging for all images, with limited fruiting specimens available for scrutiny. As there was little useful fruit or seed material collected from Singapore and Johore, physical access to the SING collections aided the discovery of a single Peninsular Malaysian collection of *Madhuca kingiana* with seeds (*Everett FRI 13942* (SING)), collected in Perak and providing a good view of the longitudinal and abaxial seed scars (Fig. 6B).

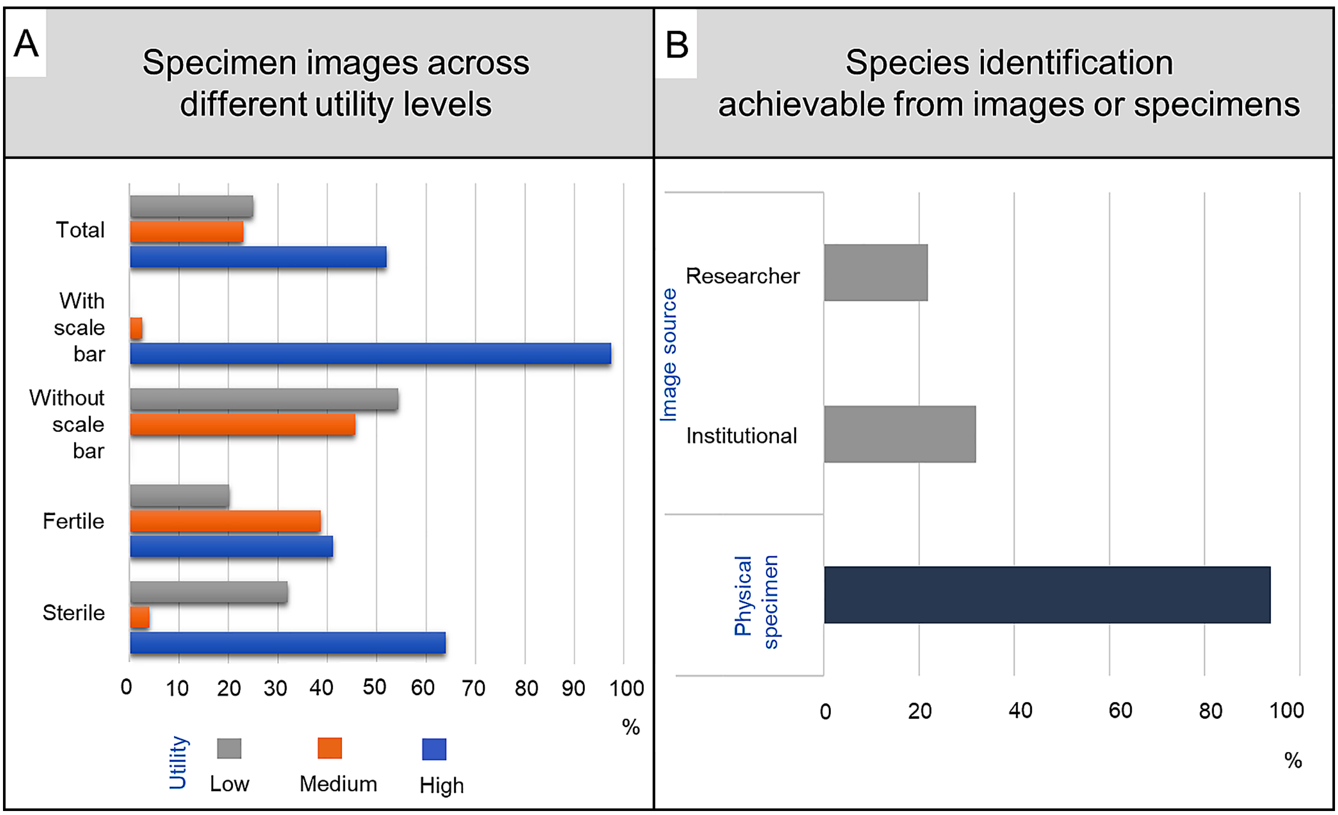

Fig. 4. **A**, Percentage of specimen images across different levels of taxonomic utility; **B**, Percentage of positive species identifications achievable from images (of different sources) and physical specimens.

■ DISCUSSION

There is no doubt that digitisation initiatives across the world have provided taxonomists access to herbarium specimens that would have otherwise been unavailable to them. This may be due to geographical location, distance from collections or resource constraints. It is also clear that digitisation has been invaluable during national COVID-19 lockdowns. These initiatives have made images and associated data available to taxonomists as well as a wide range of other scientific and non-scientific researchers including ecologists, climate scientists, biogeographers, computer modelers, pollution researchers, historians and artists (e.g., Peñuelas & Filella, 2002; Funk, 2003; IUCN, 2012; Yoshida & al., 2013; DRYFLOR, 2016; Rudin & al., 2017; Lang & al., 2018, Moonlight & al., 2020; Willis & al., 2020) as a source of information and inspiration.

Digitisation has allowed examination of specimen images and access to taxonomic data in an efficient and relatively inexpensive way. Many taxonomists today routinely use institutional images to assist with species identifications and nomenclatural clarification, such as the designation of suitable lectotypes. However, for comprehensive taxonomic accounts requiring precise species delimitation such as in this study, critical micro-morphological details such as fertile parts and indumentum were often impossible to examine from both institutional and researcher images, and physical access to specimens was needed to complete the descriptions.

It is clear that aspects of the taxonomic process can be accelerated using specimen images, in that many images can be looked at without the need to carefully handle each individual specimen separately, and that virtual analysis of images also allows pre-identification of specimens for physical dissections and closer study, thus reducing the number of characters that needed to be verified in the herbarium. This also reduces the number of specimens that need to be shipped and loaned to other institutes, which in turn protects them from damage or loss in transit, helping to maintain their availability at the home institution.

Limitations of images will of course vary between plant groups and may not be an issue for those possessing diagnostic characters that are mainly macro-morphological in scale. In these cases, a taxonomic account based on virtual specimen images alone, in theory, could be produced. However, we contend that for the large majority of angiosperm groups, access

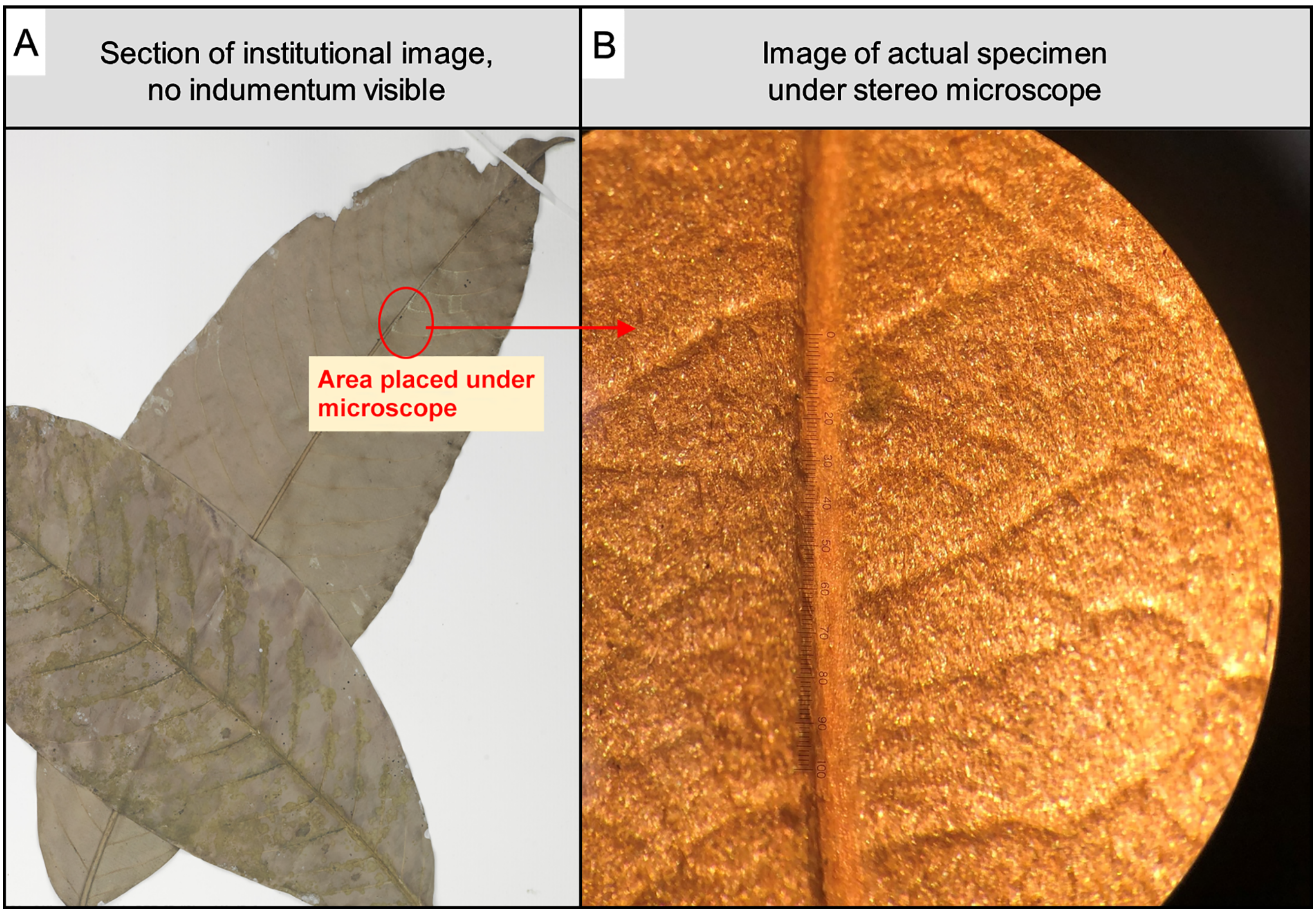

Fig. 5. Images of *Leong & al., MR 2014-049* (SING) from (**A**) institutional image and (**B**) actual specimen placed under a microscope showing sericeous indumentum.

to physical herbarium material for analysis remains critical. A survey of monographic and revisionary works for a range of genera, including trees, in the tropics and neotropics (Gibbs & Semir, 2002; Hilliard, 2004; Middleton, 2007a,b; Chantarasuwan & al., 2013; Junikka & al., 2016) highlighted that species differentiation in keys would necessitate the use of micromorphological and/or small fertile characters, including indumentum on style or ovary, androecium, anther form or seeds, amongst others also encountered in this study. This is to be expected as many families contain genera with similar macromorphological or sterile characters such as leaf arrangement (e.g., the majority of Annonaceae have distichous phyllotaxy, or spiral/opposite in Apocynaceae). A study of selected *Artocarpus* species in Thailand indicated that some diagnostic characteristics could not even be observed with a normal stereo microscope, for instance, papillae on veinlets could only be seen under a scanning electron microscope (Aneklaphakij & al., 2020).

It must be remembered that although the number of herbarium specimen images available online is increasing all the time, not all herbaria have the resources or facilities to digitise all their material. These, often smaller herbaria, still hold important specimens that need to be examined in order to undertake rigorous taxonomic research (Delves, 2021) and must not be deprioritised or ignored because images or online information is not available.

Furthermore, as shown in this study, physical checks in the herbarium can help discover useful specimens that have not yet been digitised, are filed away in the wrong place or misidentified. Some specimens may also be excluded from the digitisation process (e.g., a very degraded sample requiring specialised care, or samples in carpological and spirit collections that as far as we could tell are not routinely digitised), and it is clear that only physical access to the collections can make sure that all specimens are examined. This situation may be replicated in other herbaria where specimens are either not identified fully, or are included in backlogs.

Specialised databases such as the SRC expand the range of available information beyond major institutional catalogues, and it is likely that the virtual study of taxonomic groups without such affiliated databases may be hampered by missing specimen information. There are critically useful specimens held at many institutions yet to undertake digitisation that are inaccessible online unless visiting researchers upload records to platforms such as the SRC, thus documenting their existence to allow further scrutiny. Family- or group-specific databases, underpinned by the expertise of specialist researchers, thus help to fill information gaps outside existing institutional platforms; they capture specimen and field images from around the globe. Field images tied to a specimen are especially useful for detailed species studies, since they are more easily verifiable than field images not backed by specimen vouchers. Occurrence

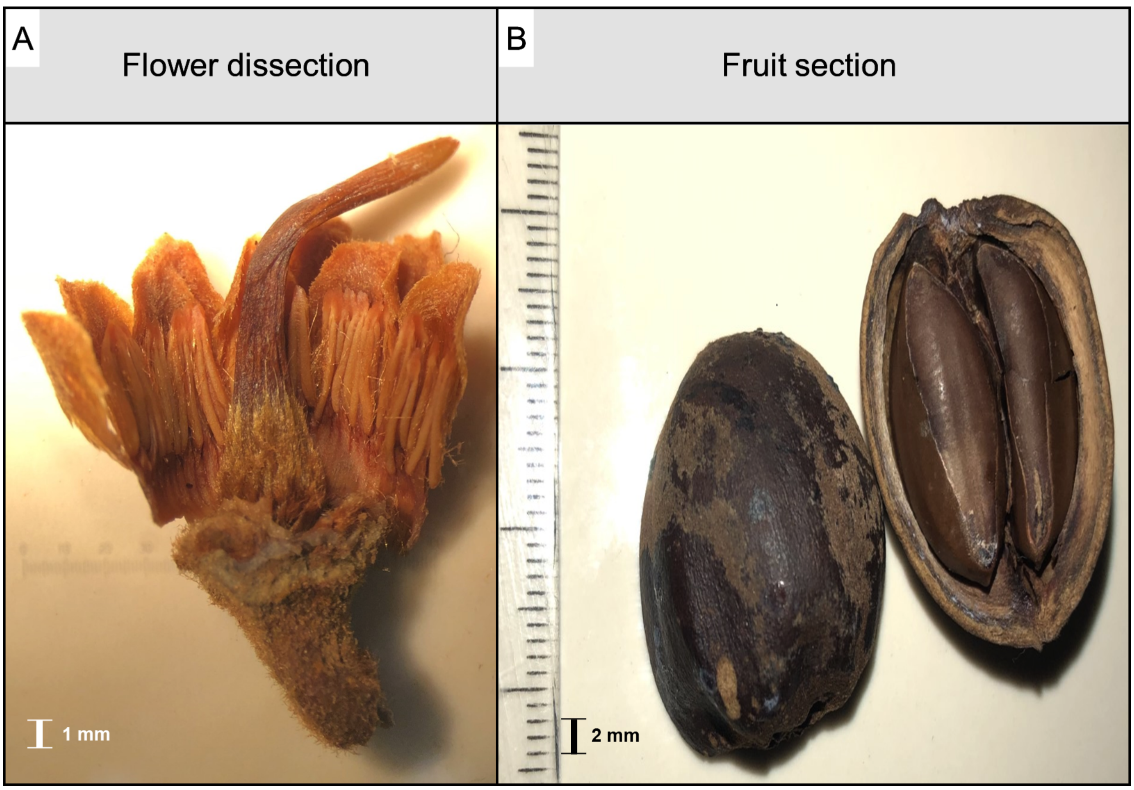

Fig. 6. A, Stereo microscope images of dissected herbarium material of *Madhuca kingiana* flower (*Ngo SING 2019-207* (SING)), allowing clear views of internal morphology and indumentum details; **B**, Fruit section of *M. kingiana* (*Everett FRI 13942* (SING)), allowing clear views of seed scars.

records in other databases that include crowdsourced data, such as GBIF, may have geographical coordinates without images, or vice versa, which make species confirmations and thus an assessment of data quality challenging. In this study, researcher images could have even greater taxonomic utility, almost matching those of institutional images, by the simple addition of a ruler or scale bar.

If herbarium access had not been possible during this study, the following would have been needed to produce a useful taxonomic account of *Madhuca* using only images:

(1) All specimen images to have scale bars included and capsules open. Not all institutional images currently capture capsule contents; it requires an additional step in the digitisation process for staff to lay out capsule contents with tweezers prior to image capture, as is the case at the SING herbarium (Fig. 7).

(2) Enhanced conditions for image contrast. Light conditions for herbarium specimen digitisation have been published (Tulig & al., 2012; Nelson & al., 2015), which could sharpen the contrast on institutional images to capture venation details, including level of indentation or protrusion. Potential enhancements include leaf clearing methods (Wilf & al., 2016), which capture chemically treated preserved leaves to show veins, or additional images created from each specimen (Unger & al., 2016, using the "Lazy Snapping" method).

(3) Microscope-level images of relevant specimens that focus on essential micro-morphological characters identified by a taxonomist familiar with the family. For Sapotaceae, these would include images of indumentum on twigs, both leaf surfaces and midrib, and all fertile parts.

(4) Dissection images (if microscope images not available) or a network of herbarium staff, curators and/or research collaborators to make them accessible virtually. Platforms such as the Digitisation Working Group in the Consortium of European Taxonomic Facilities (CETAF, 2021) may be a useful coordination tool.

The importance of enhancing digital collections to mitigate missing data on specimen images has been noted by Borges & al. (2020), who suggest capturing more images with varying magnifications and exposing hidden structures. This could be done with researchers attaching images of their own dissections to institutional images of herbarium specimens, although this requires an active feedback channel between researchers and herbaria. Targeted human input through crowdsourcing (Zhou & al., 2018) may also provide some avenues to address missing observations.

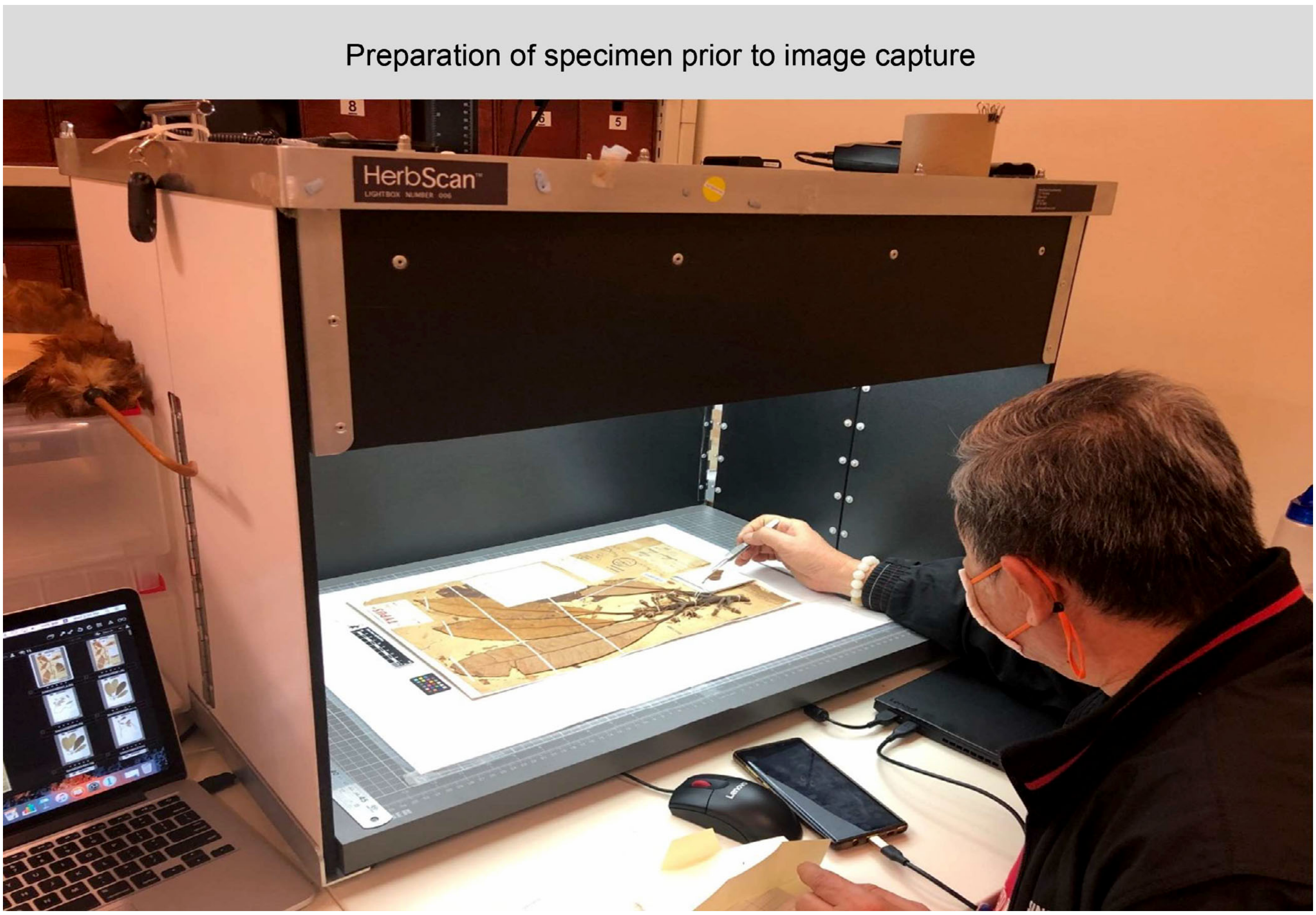

Fig. 7. Specimen scanning setup at SING herbarium, with staff laying out capsule contents with tweezers prior to image capture.

CONCLUSION

COVID-19 lockdowns have accelerated the use of digital images in plant taxonomic research. Restricted access to global herbarium collections and limited fieldwork have brought into focus the benefits and limitations of online specimen images. While these images are of significant value, physical access to specimens is still clearly needed. Diagnostic micro-morphological characters, often critical in defining species boundaries, cannot be seen in many digital images, even at the highest resolutions. We contend that these difficulties would be encountered in most tropical plant groups.

As digitisation continues apace, it is crucial that we use our pandemic experiences to develop and modify new and existing digital activities and resources to maximise their use to the taxonomic community, who are arguably the main users of the collections and who document the critical information needed to address the worldwide biodiversity crisis.

As more institutions move toward digitisation of their collections, it would be helpful for international protocols to agree to the minimum requirements necessary for individual taxonomic groups so that the main users of these images, the taxonomists, can complete species-level revisions remotely. This will of course depend upon institutional resources and ability to speedily capture images of high-enough quality (including the diagnostic characters needed) and access to the data associated with the specimen. It will also require the willingness of institutes and individuals to share specimen information either publicly or through expert researcher networks. Global working groups such as the Biodiversity Information Standards (TDWG, 2022), which seeks to establish a harmonising framework for Minimum Information about a Digital Specimen (MIDS), could potentially take a differential approach for specific taxonomic groups.

The increase in the availability of digital images has clear potential for machine learning to automate the identification of herbarium specimens and speed up the taxonomic process. However, for genera like *Madhuca* containing around a hundred species and often a relatively low number of specimens per species (some less than ten), machine learning algorithms are likely to struggle (Unger & al., 2016; Little & al., 2020). Applications such as LeafMachine (Weaver & al., 2020) function sub-optimally with cluttered leaf arrangements, poor lighting, blade overlaps/folding or extraneous fragment packets, many of the things highlighted in this paper that also limit taxonomic research using images. If some of the suggestions made in this paper are incorporated into digitisation initiatives, there could be significant benefits to researchers and their ability to speed up the documentation of plant life on Earth.

AUTHOR CONTRIBUTIONS

AP undertook the investigation, formal analysis, data curation and writing of the manuscript. PW provided conceptualisation and resources, and both PW and HA provided supervision, review and edits to the manuscript, with final acceptance by all authors. — AP, https://orcid.org/0000-0002-0615-0240; HA, https://orcid.org/0000-0002-3523-1883; PW, https://orcid.org/0000-0002-3407-8397

ACKNOWLEDGEMENTS

This paper is partly based on the dissertation by AP for her M.Sc. in Biodiversity and Plant Taxonomy at the University of Edinburgh and Royal Botanic Garden Edinburgh, supervised by PW and HA. The assistance of staff at the SBG Herbarium and RBGE Sapotaceae Resource Centre is gratefully acknowledged. All herbaria mentioned in the paper are thanked for allowing access to their collections and for allowing researchers to take photographs. The Forest Research Institute Malaysia (FRIM) is particularly thanked for supporting the last author during his time visiting Malaysian herbaria and forests under EPU permit number: 40/20019/2522. The authors would also like to thank Professor Pete Hollingsworth for helpful views on early drafts of the paper.

LITERATURE CITED

ALA 2021. Atlas of Living Australia. Website at https://www.ala.org.au/ (accessed 1 Nov 2021).

Aneklaphakij, C., Bunsupa, S., Sirichamorn, Y., Bongcheewin, B. & Satitpatipan, V. 2020. Taxonomic notes on the 'Mahat' (*Artocarpus lacucha* and *A. thailandicus*, Moraceae) species complex in Thailand. *Plants* 9(3): 391. https://doi.org/10.3390/plants9030391

Assem, J.v.d. 1953. Revision of the Sapotaceae of the Malaysian area in a wider sense. 4. *Ganua* Pierre ex Dubard. *Blumea* 7(2): 364–400.

Atrium 2021. Atrium Biodiversity Information System. Website at http://atrium.andesamazon.org (accessed 1 Nov 2021).

Baskauf, S.J. & Kirchoff, B.K. 2008. Digital plant images as specimens: Toward standards for photographing living plants. *Vulpia* 7: 16–30. https://projects.ncsu.edu/cals/plantbiology/ncsc/vulpia/pdf/Baskauf_&_Kirchoff_Digital_Plant_Images.pdf

Belbin, L., Wallis, E., Hobern, D. & Zerger, A. 2021. The Atlas of Living Australia: History, current state and future directions. *Biodivers. Data J.* 9: e65023. https://doi.org/10.3897%2FBDJ.9.e65023

Blagoderov, V., Kitching, I., Livermore, L., Simonsen, T.J. & Smith, V.S. 2012. No specimen left behind: Industrial scale digitisation of natural history collections. *ZooKeys* 209: 133–146. https://doi.org/10.3897/zookeys.209.3178

Borges, L.M., Reis, V.C. & Izbicki, R. 2020. Schrödinger's phenotypes: Herbarium specimens show two-dimensional images are both good and (not so) bad sources of morphological data. *Meth. Ecol Evol.* 11(10): 1296–1308. https://doi.org/10.1111/2041-210X.13450

Borsch, T., Stevens, A.-D., Häffner, E., Güntsch, A., Berendsohn, W.G., Appelhans, M.S., Barilaro, C., Beszteri, B., Blattner, F.R., Bossdorf, O., Dalitz, H., Dressler, S., Duque-Thüs, R., Esser, H.-J., Franzke, A., Goetze, D., Grein, M., Grünert, U., Hellwig, F., Hentschel, J., Hörandl, E., Janßen, T., Jürgens, N., Kadereit, G., Karisch, T., Koch, M.A., Müller, F., Müller, J., Ober, D., Porembski, S., Poschlod, P., Printzen, C., Röser, M., Sack, P., Schlüter, P., Schmidt, M., Schnittler, M., Scholler, M., Schultz, M., Seeber, E., Simmel, J., Stiller, M., Thiv, M., Thüs, H., Tkach, N., Triebel, D., Warnke, U., Weibulat, T., Wesche, K., Yurkov, A. & Zizka, G. 2020. A complete digitisation of German herbaria is possible, sensible and should be started now. *Res. Ideas Outcomes* 6: e50675. https://doi.org/10.3897/rio.6.e50675

Carranza-Rojas, J., Goeau, H., Bonnet, P., Mata-Montero, E. & Joly, A. 2017. Going deeper in the automated identification of herbarium specimens. *B. M. C. Evol. Biol.* 17: 181. https://doi.org/10.1186/s12862-017-1014-z
CETAF 2021. Consortium of European Taxonomic Facilities. Website at https://cetaf.org/ (accessed 1 Feb 2022).
Chantarasuwan, B., Berg, C.C. & Van Welzen, P.C. 2013. A revision of *Ficus* subsection *Urostigma* (Moraceae). *Syst. Bot.* 38(3): 653–686. https://doi.org/10.1600/036364413X670241
Chong, K.Y., Tan, H.T. & Corlett, R.T. 2009. *A checklist of the total vascular plant flora of Singapore: Native, naturalised and cultivated species*. Singapore: Raffles Museum of Biodiversity Research.
Corney, D.P.A., Clark, J.Y., Tang, H.L. & Wilkin, P. 2012. Automatic extraction of leaf characters from herbarium specimens. *Taxon* 61(1): 231–244. https://doi.org/10.1002/tax.611016
CRIA [Centro de Referência em Informação Ambiental] 2021. *species*Link. Website at https://specieslink.net (accessed 1 Nov 2021).
Davis, C.C., Kennedy, J.A. & Grassa, C.J. 2021. Back to the future: A refined single-user photostation for massively scaling herbarium digitisation. *Taxon* 70(3): 635–643. https://doi.org/10.1002/tax.12459
Delves, J. 2021. *Decolonise herbaria and specimen data: Quantifying the contribution of local herbaria to biodiversity patterns*. M.Sc. thesis. University of Edinburgh and Royal Botanic Garden Edinburgh, Scotland, U.K.
DiSSCo 2021. Distributed System of Scientific Collections. Website at https://www.dissco.eu/ (accessed 1 Nov 2021).
Drinkwater, R.E., 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%2Fphytokeys.38.7168
DRYFLOR 2016. Plant diversity patterns in neotropical dry forests and their conservation implications. *Science* 353(6306): 1383–1387. https://doi.org/10.1126/science.aaf5080
Flora of Nepal 2014–. floraofnepal.org. Royal Botanic Garden Edinburgh. Website at http://www.floraofnepal.org (accessed 1 Nov 2021).
Flora of Singapore 2019–. Singapore: National Parks Board. Website at https://www.nparks.gov.sg/sbg/research/publications/flora-of-singapore (accessed 1 Nov 2021).
Funk, V.A. 2003. 100 uses for an herbarium (Well at least 72). *Amer. Soc. Pl. Taxon. Newslett.* 17(2): 17–19. https://repository.si.edu/bitstream/handle/10088/11385/bot_2003_pr_Funk_100Uses.pdf
GBIF 2021. Global Biodiversity Information Facility. Website at https://www.gbif.org/ (accessed 1 Nov 2021).
Gibbs, P. & Semir, J. 2002. A taxonomic revision of the genus *Ceiba* Mill. (Bombacaceae). *Anales Jard. Bot. Madrid* 60(2): 259–300.
Häuser, C.L., Steiner, A., Holstein, J. & Scoble, M.J. (eds.) 2005. *Digital imaging of biological type specimens: A manual of best practice; Results from a study of the European Network for Biodiversity Information*. Stuttgart: Staatliches Museum für Naturkunde. https://researchportal.unamur.be/files/12004910/enbi_ImagingBiologicalSpecimens_manual_en_v1.pdf
Hiepko, P. (ed.) 1972. *Herbarium Willdenow alphabetical index*. Zug: Inter Documentation Company AG.
Hilliard, O.M. 2004. A revision of *Chirita* sect. *Liebigia* (Gesneriaceae). *Edinburgh J. Bot.* 60(3): 361–387. https://doi.org/10.1017/S0960428603000295
Holovachov, O., Zatushevsky, A. & Shydlovsky, I. 2014. Whole-drawer imaging of entomological collections: Benefits, limitations and alternative applications. *J. Conservation Mus. Stud.* 12(1): 9. https://doi.org/10.5334/jcms.1021218
IDC [undated]. The Wallich Herbarium (the herbarium of the East Indian Company) on microfiche. Zug: Inter Documentation Company AG.
iDigBio 2021. Integrated Digitised Biocollections. Website at https://www.idigbio.org/ (accessed 1 Nov 2021).
IPNI 2021. The International Plant Names Index. Published on the Internet. The Royal Botanic Gardens, Kew, Harvard University Herbaria & Libraries and Australian National Botanic Gardens. https://www.ipni.org/ (accessed 1 Nov 2021).
IUCN 2012. *IUCN red list categories and criteria: Version 3.1*, 2nd ed. Gland & Cambridge: IUCN.
JSTOR 2018a. *JSTOR Global Plants: Guidelines for scanning specimens*. Pdf at https://guides.jstor.org/ld.php?content_id=31764146 (accessed 1 Nov 2021).
JSTOR 2018b. *JSTOR Plants Handbook*. http://www.snsb.info/SNSBInfoOpenWiki/attach/Attachments/JSTOR-Plants-Handbook.pdf (accessed 1 Nov 2021).
Junikka, L., Maas, P.J.M., Maas-Van De Kamer, H. & Westra, L.Y.T. 2016. Revision of *Oxandra* (Annonaceae). *Blumea* 61(3): 215–266. https://doi.org/10.3767/000651916X694283
King, S., Pinon, J. & Drinkwater, R. 2019. Utilising the crowd to unlock the data on herbarium specimens at the Royal Botanic Garden Edinburgh. *Biodivers. Inf. Sci. Stand.* 12. https://link.gale.com/apps/doc/A646406344/AONE?u=anon~c4e8d066&sid=googleScholar&xid=56fb528c
LaFrankie, J.V. 1996. Distribution and abundance of Malayan trees: Significance of family characteristics for conservation. *Gard. Bull. Singapore* 48: 75–87.
Lang, P.L.M., Willems, F.M., Scheepens, J.F., Burbano, H.A. & Bossdorf, O. 2018. Using herbaria to study global environmental change. *New Phytol.* 221(1): 110–122. https://doi.org/10.1111/nph.15401
Le Bras, G., Pignal, M., Jeanson, M., Muller, S., Aupic, C., Carré, B., Flament, G., Gaudeul, M., Gonçalves, C., Invernón, V., Jabbour, F., Lerat, E., Lowry, 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. *Sci. Data* 4: 170016. https://doi.org/10.1038/sdata.2017.16
Lendemer, J., Thiers, B., Monfils, A.K., Zaspel, J., Ellwood, E.R., Bentley, A., LeVan, K., Bates, J., Jennings, D. & Contreras, D. 2020. The extended specimen network: A strategy to enhance US biodiversity collections, promote research and education. *Bioscience* 70(1): 23–30. https://doi.org/10.1093/biosci/biz140
Little, D.P., Tulig, M., Tan, K.C., Liu, Y., Belongie, S., Kaeser-Chen, C., Michelangeli, F.A., Panesar, K., Guha, R.V. & Ambrose, B.A. 2020. An algorithm competition for automatic species identification from herbarium specimens. *Applic. Pl. Sci.* 8(6): e11365. https://doi.org/10.1002/aps3.11365
Middleton, D.J. 2007a. *Flora Malesiana*, ser. 1, *Seed plants*, vol. 18, *Apocynaceae (subfamilies Rauvolfioideae and Apocynoideae)*. Leiden: Nationaal Herbarium Nederland.
Middleton, D.J. 2007b. A revision of *Aeschynanthus* (Gesneriaceae) in Thailand. *Edinburgh J. Bot.* 64(3): 363–429. https://doi.org/10.1017/S0960428607004878
Middleton, D.J., Armstrong, K., Baba, Y., Balslev, H., Chayamarit, K., Chung, R.C.K., Conn, B.J., Fernando, E.S., Fujikawa, K., Kiew, R. & Luu, H.T. 2019. Progress on Southeast Asia's flora projects. *Gard. Bull. Singapore* 71(2): 267–319. https://doi.org/10.26492/gbs71(2).2019-02
Moonlight, P.W., Silva de Miranda, P.L., Domingos Cardoso, D., Dexter, K.G., Oliveira-Filho, A.T., Pennington, R.T., Ramos, R. & Särkinen, T.E. 2020. The strengths and weaknesses of species distribution models in biome delimitation. *Global Ecol.* 29(10): 1770–1784. https://doi.org/10.1111/geb.13149
Nelson, G., Sweeney, P., Wallace, L.E., Rabeler, R.K., Allard, D., Brown, H., Carter, J.R., Denslow, M.W., Ellwood, E.R. & Germain-Aubrey, C.C. 2015. Digitisation workflows for flat sheets and packets of plants, algae, and fungi. *Applic. Pl. Sci.* 3(9): 1500065. https://doi.org/10.3732/apps.1500065
Ng, F. 1972. Sapotaceae. Pp. 388–439 in: Whitmore, T.C. (ed.), *Tree flora of Malaya*, vol. 1. Kuala Lumpur: Longman Malaysia.
Nieva de la Hidalga, A., Rosin, P.L., Sun, X., Bogaerts, A., De Meeter, N., De Smedt, S., Strack van Schijndel, M., Van

Wambeke, P. & Groom, Q. 2020. Designing an herbarium digitisation workflow with built-in image quality management. *Biodivers. Data J.* 8: e47051. https://doi.org/10.3897%2FBDJ.8.e47051
Pennington, T.D. 1991. *The genera of Sapotaceae*. London: Royal Botanic Gardens Kew.
Peñuelas, J. & Filella, I. 2002. Metal pollution in Spanish terrestrial ecosystems during the twentieth century. *Chemosphere* 46(4): 501–505. https://doi.org/10.1016/S0045-6535(01)00171-0
Pryer, K.M., Tomasi, C., Wang, X., Meineke, E.K. & Windham, M.D. 2020. Using computer vision on herbarium specimen images to discriminate among closely related horsetails (*Equisetum*). *Applic. Pl. Sci.* 8(6): e11372. https://doi.org/10.1002/aps3.11372
RBGE 2021. Royal Botanic Garden: Searchable Resource Centres. Facilitated by the Royal Botanic Garden Edinburgh. Available at https://www.rbge.org.uk/collections/online-resources (accessed 1 Nov 2021).
Reflora – Virtual Herbarium 2021. Available at http://floradobrasil.jbrj.gov.br/reflora/herbarioVirtual/ (accessed 1 Nov 2021).
Royen, P. van 1960. Revision of the Sapotaceae of the Malaysian area in a wider sense. 20. *Madhuca* Gmelin. *Blumea* 10(1): 1–117.
Rudin, S.M., Murray, D.W. & Whitfeld, T.J.S. 2017. Retrospective analysis of heavy metal contamination in Rhode Island based on old and new herbarium specimens. *Applic. Pl Sci.* 5(1): 1600108. https://doi.org/10.3732/apps.1600108
Slik, J.F., Arroyo-Rodríguez, V., Aiba, S.-I., Alvarez-Loayza, P., Alves, L.F., Ashton, P., Balvanera, P., Bastian, M.L., Bellingham, P.J., Berg, E.v.d., Bernacci, L., Bispo, P.d.C., Blanc, L., Böhning-Gaese, K., Boeckx, P., Bongers, F., Boyle, B., Bradford, M., Brearley, F.Q., Hockemba, M.B.-N., Bunyavejchewin, S., Matos, D.C.L., Castillo-Santiago, M., Catharino, E.L.M., Chai, S.-L., Chen, Y., Colwell, R.K., Chazdon, R.L., Clark, C., Clark, D.B., Clark, D.A., Culmsee, H., Damas, K., Dattaraja, H.S., Dauby, G., Davidar, P., DeWalt, S.J., Doucet, J.-L., Duque, A., Durigan, G., Eichhorn, K.A.O., Eisenlohr, P.V., Eler, E., Ewango, C., Farwig, N., Feeley, K.J., Ferreira, L., Field, R., Filho, A.T.d.O., Fletcher, C., Forshed, O., Franco, G., Fredriksson, G., Gillespie, T., Gillet, J.-F., Amarnath, G., Griffith, D.M., Grogan, J., Gunatilleke, N., Harris, D., Harrison, R., Hector, A., Homeier, J., Imai, N., Itoh, A., Jansen, P.A., Joly, C.A., Jong, B.H.J.d., Kartawinata, K., Kearsley, E., Kelly, D.L., Kenfack, D., Kessler, M., Kitayama, K., Kooyman, R., Larney, E., Laumonier, Y., Laurance, S., Laurance, W.F., Lawes, M.J., Amaral, I.L.d., Letcher, S.G., Lindsell, J., Lu, X., Mansor, A., Marjokorpi, A., Martin, E.H., Meilby, H., Melo, F.P.L., Metcalfe, D.J., Medjibe, V.P., Metzger, J.P., Millet, J., Mohandass, D., Montero, J.C., Valeriano, M.d.M., Mugerwa, B., Nagamasu, H., Nilus, R., Ochoa-Gaona, S., Onrizal, Page, N., Parolin, P., Parren, M., Parthasarathy, N., Paudel, E., Permana, A., Piedade, M.T.F., Pitman, N.C.A., Poorter, L., Poulsen, A.D., Poulsen, J., Powers, J., Prasad, R.C., Puyravaud, J.-P., Razafimahaimodison, J.-C., Reitsma, J., Santos, J.R.d., Spironello, W.R., Romero-Saltos, H., Rovero, F., Rozak, A.H., Ruokolainen, K., Rutishauser, E., Saiter, F., Saner, P., Santos, B.A., Santos, F., Sarker, S.K., Satdichanh, M., Schmitt, C.B., Schöngart, J., Schulze, M., Suganuma, M.S., Sheil, D., Pinheiro, E.d.S., Sist, P., Stevart, T., Sukumar, R., Sun, I.-F., Sunderland, T., Suresh, H.S., Suzuki, E., Tabarelli, M., Tang, J., Targhetta, N., Theilade, I., Thomas, D.W., Tchouto, P., Hurtado, J., Valencia, R., Valkenburg, J.L.C.H.v., Do, T.V., Vasquez, R., Verbeeck, H., Adekunle, V., Vieira, S.A., Webb, C.O., Whitfeld, T., Wich, S.A., Williams, J., Wittmann, F., Wöll, H., Yang, X., Yao, C.Y.A., Yap, S.L., Yoneda, T., Zahawi, R.A., Zakaria, R., Zang, R., Assis, R.L.d., Luize, B.G. & Venticinque, E.M. 2015. An estimate of the number of tropical tree species. *Proc. Natl. Acad. Sci. U.S.A.* 112(24): 7472–7477. https://doi.org/10.1073/pnas.1423147112
Solanaceae Source 2021. Solanaceae Source: A global taxonomic resource for the nightshade family. Website at http://solanaceaesource.org/ (accessed 1 Nov 2021).
Takano, A., Horiuchi, Y., Fujimoto, Y., Aoki, K., Mitsuhashi, H. & Takahashi, A. 2019. Simple but long-lasting: A specimen imaging method applicable for small- and medium-sized herbaria. *PhytoKeys* 118: 1–14. https://doi.org/10.3897/phytokeys.118.29434
TDWG 2022. Biodiversity Information Standards. https://tdwg.org (accessed 1 Mar 2022).
Thiers, B. 2021. Index Herbariorum: A global directory of public herbaria and associated staff. *New York Botanical Garden. Available at* http://sweetgum.nybg.org/science/ih/ (accessed 1 Nov 2021).
Tulig, M., Tarnowsky, N., Bevans, M., Kirchgessner, A. & Thiers, B.M. 2012. Increasing the efficiency of digitisation workflows for herbarium specimens. *Zookeys* 209: 103–113. https://doi.org/10.3897/zookeys.209.3125
Turner, I.M. 1995. A catalogue of the vascular plants of Malaya. *Gard. Bull. Singapore* 47(1–2): 1–757.
Unger, J., Merhof, D. & Renner, S. 2016. Computer vision applied to herbarium specimens of German trees: Testing the future utility of the millions of herbarium specimen images for automated identification. *B. M. C. Evol. Biol.* 16: 248. https://doi.org/10.1186/s12862-016-0827-5
WCSP 2021. World Checklist of Selected Plant Families. Facilitated by the Royal Botanic Gardens, Kew. Published on the internet at https://wcsp.science.kew.org/ (accessed 1 Nov 2021).
Weaver, W.N., Ng, J. & Laport, R.G. 2020. LeafMachine: Using machine learning to automate leaf trait extraction from digitised herbarium specimens. *Applic. Pl. Sci.* 8(6): e11367. https://doi.org/10.1002/aps3.11367
Wilf, P., Zhang, S., Chikkerur, S., Little, S.A., Wing, S.L. & Serre, T. 2016. Computer vision cracks the leaf code. *Proc. Natl. Acad. Sci. U.S.A.* 113(12): 3305–3310. https://doi.org/10.1073/pnas.1524473113
Wilkie, P., Pullan, M., Richardson, J., Armstrong, K., Buendia, J., Culverhouse, F., Elliott, A., Fabriani, F., Kumarage, L., Mackinder, B., Purvis, D., Solander, S. & Woods, E. 2008–. Sapotaceae Resource Centre. Online Database. http://padme.rbge.org.uk/sapotaceae (accessed 1 Nov 2021).
Willis, 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. 2020. Old plants, new tricks: Phenological research using herbarium specimens. *Trends Ecol. Evol.* 32(7): 531–546. https://doi.org/10.1016/j.tree.2017.03.015
WFO 2021. World Flora Online. Published on the internet at http://www.worldfloraonline.org (accessed 1 Nov 2021).
Yoshida, K., Schuenemann, V.J., Cano, L.M., Pais, M., Mishra, B., Sharma, R., Lanz, C., Martin, F.N., Kamoun, S., Krause, J., Thines, M., Weigel, D. & Burbano, H.A. 2013. The rise and fall of the *Phytophthora infestans* lineage that triggered the Irish potato famine. *eLife* 2: e00731. https://doi.org/10.7554/eLife.00731.001
Zhou, N., Siegel, Z.D., Zarecor, S., Lee, N., Campbell, D.A., Andorf, C.M., Nettleton, D., Lawrence-Dill, C.J., Ganapathysubramanian, B., Kelly, J.W. & Friedberg, I. 2018. Crowdsourcing image analysis for plant phenomics to generate ground truth data for machine learning. *PLoS Computat. Biol.* 14(7): e1006337. https://doi.org/10.1371/journal.pcbi.1006337